An Investigation into the Sugar-substrate Specificity of the Sialic Acid Synthases

---

A thesis submitted for the fulfilment of the

requirements for the degree of

Master of Science in Biochemistry

at the

University of Canterbury

by

Thomas Richard Cotton

Department of Chemistry

University of Canterbury

September 2014

Acknowledgements

First and foremost, thank you Emily Parker for your continued support as my supervisor. You have made my time at Canterbury extremely rewarding and I've learnt a great deal from you. Thank you for being constantly available for advice/discussions and for encouraging me to push myself. Without your support I certainly wouldn't be in the position I am now, and I am truly grateful for the massive contribution you have made to my education.

I extend this thanks to the rest of the Parker group, past and present, for providing invaluable assistance in all aspects of my research. Dr. Dmitri Joseph, I owe you big time. From teaching me almost everything I know in the lab, to providing ongoing support throughout my project, you've been a massive help and I wish you all the best for whatever you go on to do. Thanks to Ryu Toyama, the third sialic acid synthase musketeer, for all your help with lab work in the early days, it certainly made my life a lot easier.

Thank you to Penel Cross, Gerd Mittelstaedt, Nicky Blackmore, Ali Reza Nazmi, Wanting Jiao, and Sarah Wilson-Coutts for contributing your valuable time and expertise, inside the lab and out. It's been a great pleasure working with you all, and I'm forever grateful for everything you've taught me. GJ and Logan, cheers for the help with all things chemistry, the excessive daily coffee breaks, the after work beers, the quality chat and extracurricular activities. It's been great getting to know you guys. Eric Lang, it's been a pleasure sharing this side of the office with you and thanks for providing me with some insight into the fascinating world of computational chemistry. Vicky and Effie, thanks for your friendship and help over the years. I'm lucky to have shared this experience with such great people.

I would also like to thank the Chemistry Department's technical staff for all your work behind the scenes, making sure everything runs smoothly and ensuring the resources we need for this type of research are always available.

Abstract

The sialic acids represent a structurally and functionally diverse family of nine-carbon keto sugars. These compounds, derived structurally from neuraminic acid, are central to many of life's processes at the molecular level. *N*-Acetylneuraminic acid (NeuNAc) is the most naturally abundant of the sialic acids and acts as the terminal residue of mammalian cell surface glycoconjugates. The negative charge and exposed location of NeuNAc makes this compound intrinsically linked to cellular signalling, recognition and adhesion events. While NeuNAc expression is typically restricted to eukaryotic phyla, some pathogenic bacteria have evolved to express NeuNAc on their own cell surfaces, allowing them to mimic the surface physiology of their mammalian host's cells and evade detection by the immune system. In bacteria, NeuNAc biosynthesis proceeds via the aldol-like condensation reaction of phosphoenolpyruvate (PEP) with the sugar-substrate *N*-acetylmannosamine (ManNAc). Alternatively, in mammalian systems ManNAc is first phosphorylated to ManNAc 6-phosphate (ManNAc 6-P) before undergoing condensation with PEP to give NeuNAc 9-phosphate (NeuNAc 9-P). These PEP condensation reactions are catalysed by an evolutionarily related family of homo-dimeric $(\beta\alpha)_8$ barrel enzymes (NeuNAc synthase in bacteria and NeuNAc 9-P synthase in mammals), referred to collectively as the sialic acid synthases.

In addition to NeuNAc, several pathogenic bacteria synthesise a number of unique 'bacterial sialic acids', including legionaminic and pseudaminic acid, which are known to be essential for the motility and pathogenicity of some species. These alternate sialic acids are again biosynthesised via condensation reactions of PEP with variable sugar substrates, catalysed by additional members of the sialic acid synthase family.

In chapter two of this thesis, I report a structural and functional characterisation of human NeuNAc 9-P synthase. Modelling and mutagenesis were used to delineate possible sugar-substrate binding modes, with a number of potentially important phosphate binding residues identified. Biophysical analysis reveals the human enzyme adopts a domain-swapped homo-dimeric conformation in solution, as previously observed for the analogous enzyme from *Neisseria meningitidis.*

Chapter three details an investigation into the variable sugar-substrate specificity of mammalian and bacterial sialic acid synthases, which are apparently selective for their

respective sugar-substrates depending entirely upon the presence or absence of the C-6 phosphate group. Bioinformatic analysis of bacterial and mammalian sequences revealed the $\beta_2\alpha_2$ loop of the catalytic barrel as a putative sugar-substrate selectivity element. Substitution of the bacterial loop with the mammalian loop sequence however was alone insufficient to confer novel activity with the mammalian sugar-substrate.

In chapter four, the $\beta_2\alpha_2$ loop of the bacterial sialic acid synthases; legionaminic acid synthase (LegS) and pseudaminic acid synthase (PseS), was again identified as a structural element potentially involved in the sugar-substrate selectivity of these enzymes. Generation of chimeric LegS and PseS variants incorporating the $\beta_2\alpha_2$ loop from bacterial NeuNAc synthase was unsuccessful at conferring novel NeuNAc synthase activity due to unforeseen disruption of PEP binding and quaternary structure.

The work presented in this thesis provides a starting point from which to pursue a comprehensive understanding of the molecular basis of sugar-substrate specificity of the sialic acid synthases. An appreciation of how these enzymes achieve precise sugar-substrate specificity may provide the basis for their exploitation as therapeutic or biosynthetic targets.

Table of contents

List of abbreviations

AFPL	Antifreeze protein-like
Amp	Ampicillin
AUC	Analytical ultracentrifugation
BTP	Bis-tris propane
Cam	Chloramphenicol
CD	Circular Dichroism
Cje	*Campylobacter jejuni*
CMP	Cytidine-monophosphate
CMP-NeuNAc	Cytidine-monophosphate *N*-acetylneuraminic acid
CNS	Central nervous system
Da	Dalton
DAH7PS	3-Deoxy-D-*arabino*-heptulosonate 7-phosphate synthase
DMSO	Dimethyl sulfoxide
DSF	Differential scanning fluorimetry
GDP	Guanosine-diphosphate
GlcNAc	*N*-Acetylglucosamine
Hsa	*Homo sapiens*
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
kcal	Kilocalorie
k_{cat}	Turnover number

k_{cat}/K_M	Specificity constant
K_d	Dissociation constant
kDa	Kilodalton
KDN	2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid
KDN 9-P	2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid 9-phosphate
KDO8PS	2-keto-3-deoxy-D-*manno*-octulosonate-8-phosphate synthase
K_M	Michaelis-Menten constant
LB	Lysogeny broth
Leg	Legionaminic acid
LegS	Legionaminic acid synthase
Lpn	*Legionella pneumophila*
LPS	Lipopolysaccharide
Man 6-P	Mannose 6-phosphate
ManNAc	*N*-Acetylmannosamine
ManNAc 6-P	*N*-Acetylmannosamine 6-phosphate
MRSA	Methicillin resistant *Staphylococcus aureus*
MWCO	Molecular weight cut-off
NADH	Nicotinamide adenine dinucleotide
NCAM	Neuronal cell-adhesion molecule
NDP	Nucleotide-diphosphate
NeuNAc	*N*-Acetylneuraminic acid
NeuNAcS	*N*-Acetylneuraminic acid synthase
NeuNAc 9-P	*N*-Acetylneuraminic acid 9-phosphate

NeuNAc 9-PS	*N*-Acetylneuraminic acid 9-phosphate synthase
Nme	*Neisseria meningitidis*
NMR	Nuclear magnetic resonance
OD	Optical density
PCR	Polymerase chain-reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEP	Phosphoenolpyruvate
Pse	Pseudaminic acid
PseS	Pseudaminic acid synthase
rManNAc	Reduced *N*-acetylmannosamine
RMSD	Root-mean-square deviation
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
Spec	Spectinomycin
TAE	Tris-acetate-ethylenediaminetetraacetic acid
TBA	Thiobarbituric acid
TE	Tris-ethylenediaminetetraacetic acid
TEV	Tobacco etch virus
UDP	Uridine-diphosphate
UDP-GlcNAc	Uridine-diphosphate *N*-acetylglucosamine

Chapter 1: Introduction

1.1 The sialic acids

1.1.1 Overview

The sialic acids comprise a diverse family of nine-carbon α-keto sugars, the majority of which are structural derivatives of neuraminic acid (Figure 1.1).[1] Neuraminic acid has an amino moiety at the C-5 position, and carboxylate at C-2 which confers negative charge at physiological pH.[2] While unsubstituted neuraminic acid is not known to exist in nature, over 40 substituted variations (classified as sialic acids) have now been described.[2] Common modifications to the core neuraminic acid framework include *O*-acetylation and esterification of the hydroxyl moieties, namely at positions C-7, C-8 and C-9 (Figure 1.1).[3] Multiple substitutions at these positions allow further diversification of sialic acid structure. Other modifications include glycolation (e.g. *N*-glycolylneuraminic acid) and deamination (e.g. 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid) of the C-5 amino group.[3] *N*-Acetylation at position C-5 yields *N*-acetylneuraminic acid (NeuNAc), which is by far the most prevalent sialic acid in nature and has become synonymous with the term sialic acid (which more correctly refers to the collective family of compounds structurally derived from the nine-carbon neuraminic acid framework). [2]

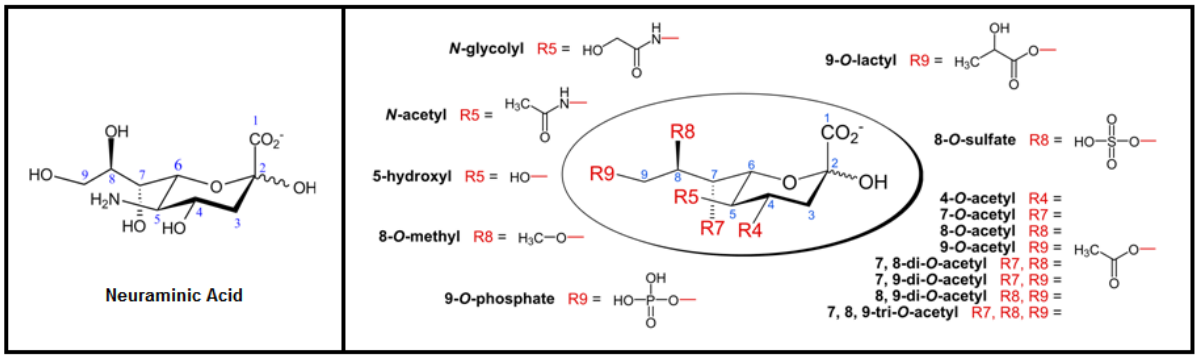


Figure 1.1. The family of naturally occurring sialic acids, structurally derived from neuraminic acid (shown left). Figure adapted from Schauer (2004).[3]

The structural diversity of the sialic acids allows these molecules to facilitate a wide range of biological functions, however sialic acid expression is not ubiquitous in nature.[2, 4] Sialic acids are commonly expressed in members of the animal kingdom, particularly in higher organisms (echinoderms upwards). NeuNAc is prevalent in mammalian systems, where it acts as the

terminal residue of cell surface glycoconjugates and thus plays central roles in cell signalling, recognition and adhesion events.[1-2, 4]

In addition to mammalian systems, biosynthesis of NeuNAc has been detected in insects, including larval stage *Drosophila melanogaster*.[5] Unlike animals, plants and prokaryotes typically lack NeuNAc expression. A definitive NeuNAc biosynthetic pathway is yet to be elucidated for any plant species however there is some evidence to suggest such pathways exist.[6] Prokaryotic expression of sialic acids is uncommon, however there are exceptions, and a number of gram-negative bacteria synthesise certain sialic acids.[7] By integrating NeuNAc into their own cell surface glycoconjugates, pathogenic bacteria are able to effectively mimic the surface physiology of their mammalian host's cell and evade detection by the immune system.[1, 7] Additionally, expression of the "bacterial sialic acids"; 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid (pseudaminic acid) and 5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid (legionaminic acid) has been reported for a number of gram-negative bacteria.[8-10] These compounds are apparently unique to bacteria and microorganisms, and are essential for the development of cell surface structures which aid in motility and pathogenicity.[9, 11]

1.1.2 N-Acetylneuraminic acid

NeuNAc contains an *N*-acetyl functionality at position C-5 (Figure 1.2) and is the most abundant sialic acid in nature.[2] In mammalian systems, NeuNAc is incorporated as the terminal residue of glycoconjugates which decorate the cell's surface.[4] Because of this exposed terminal position, and the molecule's negative charge at physiological pH, NeuNAc is involved in cellular recognition and adhesion events, notably during immune and inflammatory responses.[2-3] Sialic acid binding lectins presented on the surface of immune effector cells mediate immune-responses via interaction with sialic acid ligands expressed by other cells.[12]

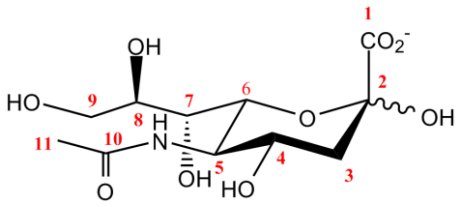


Figure 1.2. *N*-Acetylneuraminic acid is a nine-carbon amino-sugar. The C-1 carboxylate confers negative charge at physiological pH.

Under some circumstances, NeuNAc is implicated in inter-cellular repulsion via concentration of negative charge at the cell surface. This repulsive effect is thought to contribute to the metastatic potential of certain cancer cells, driving separation of cells from the bulk of the tumor mass.[13-14] The negative charge of NeuNAc may also function to mask certain recognition sites, and protect underlying glycoconjugate regions from cellular degradation.[2]

Sialic acids are expressed in nearly all mammalian tissue types, however the highest concentrations are found in the central nervous system (CNS).[15] Gangliosides, being lipid based glycoconjugates, contain terminal NeuNAc residues in their oligosaccharide portion. These gangliosides are important for brain development and neuronal function.[16] Increased NeuNAc concentration in brain gangliosides has been associated with enhanced neuritic growth, and neuro-protection from glutamate toxicity.[17] Polymeric NeuNAc is also an important constituent carbohydrate of the CNS glycoprotein, neuronal cell adhesion molecule (NCAM), which is known to be important for cellular adhesion processes pertaining to cognition and neural plasticity.[18] Proper sialylation of NCAM is required for its function, and has been shown to be essential for the foetal development of mice.[19]

Because of the abundance of exposed NeuNAc in the CNS, neuroinvasive pathogenic bacteria such as *Neisseria meningitidis* and *Escherichia coli* K1 express NeuNAc on their own cell surfaces, allowing them to mimic the exterior of mammalian cells, and thus appear immunologically silent within their hosts.[7, 20] This molecular mimicry makes it difficult for the immune system to identify and clear pathogenic invaders, and has hindered the development of polysaccharide vaccines against such pathogens.[21]

NeuNAc presented on mammalian cell surfaces also facilitates the recognition and binding of certain viruses to host cells prior to internalisation.[22] For this reason, structural analogues of NeuNAc have applications as anti-viral therapeutics, including the frontline influenza drugs Zanamavir (Relenza™) and Oseltamivir (Tamiflu®).[23] These NeuNAc analogues are potent inhibitors of the influenza neuraminidase enzyme which functions to hydrolyse the terminal NeuNAc residue from cell surface glycoconjugates.[24] By cleaving this terminal sugar moiety, nascent viral particles are able to be released from the surface of the infected cell.[24] Neuraminidase inhibitors prevent cleavage of the NeuNAc residue which binds the virus particle to the cell surface, thus preventing distribution of the virus.[25]

1.1.3 N-Acetylneuraminic acid biosynthesis

The biosynthetic pathways of NeuNAc from both mammalian and bacterial systems are well documented (Figure 1.3).[26-31] Both pathways begin with the activated sugar-substrate uridine diphosphate (UDP) *N*-acetylglucosamine (UDP-GlcNAc). Epimerisation and hydrolysis of this metabolite, catalysed by the enzyme UDP-GlcNAc 2-epimerase, generates the amino sugar *N*-acetyl mannosamine (ManNAc).[30] The bacterial pathway utilises ManNAc directly in an aldol-like condensation reaction with phosphoenolpyruvate (PEP) to form NeuNAc, a reaction catalysed by the enzyme NeuNAc synthase (NeuNAcS).[28, 32] Interestingly, the mammalian pathway involves an additional step in which ManNAc is first phosphorylated via ManNAc kinase activity to ManNAc 6-phosphate (ManNAc 6-P).[29]

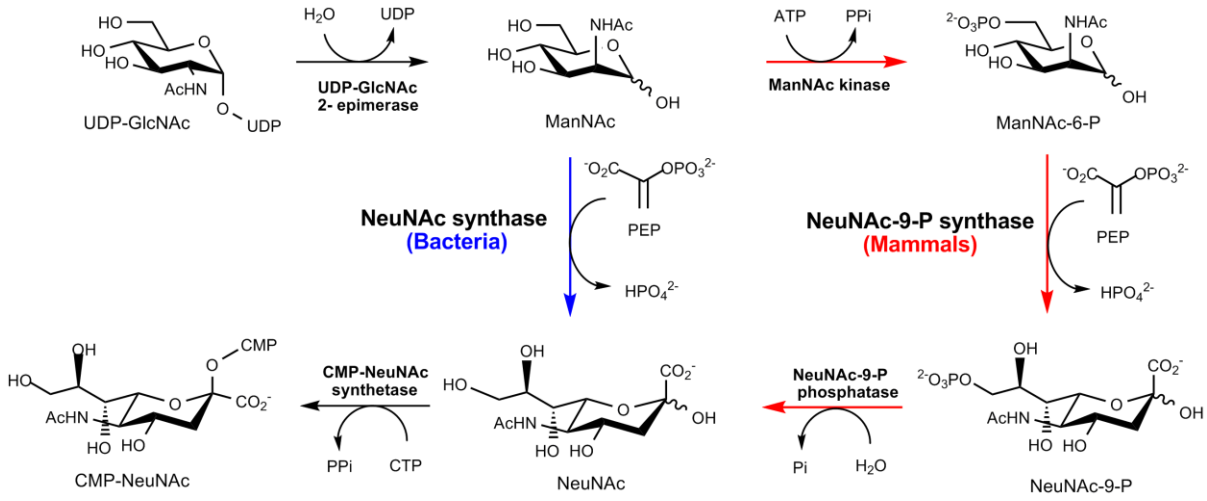


Figure 1.3. The NeuNAc biosynthetic pathway from bacterial and mammalian systems. The bacterial pathway utilises ManNAc directly for condensation with PEP to yield NeuNAc (blue arrow). The mammalian pathway includes an additional two steps in which the precursor metabolite ManNAc is phosphorylated at C-6, and then following condensation with PEP, the product NeuNAc 9-P is dephosphorylated to yield NeuNAc (red arrows).

ManNAc 6-P is the substrate for mammalian NeuNAc 9-phosphate synthase (NeuNAc 9-PS) which catalyses the aldol-like condensation of PEP with ManNAc 6-P to yield NeuNAc 9-phosphate.[4, 27, 33] This step is analogous to that catalysed by NeuNAcS in bacterial systems.[4] NeuNAc 9-P is subsequently dephosphorylated to form NeuNAc.[1] A final step in both bacterial and mammalian systems generates 'activated' NeuNAc via conjugation to cytidine-monophosphate (CMP), which acts as the glycosyl donor for incorporation into cell surface glycoconjugates.[31, 34]

1.1.4 2-Keto-3-deoxy-D-glycero-D-galacto-nononic acid

Deaminated neuraminic acid 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid (KDN) is another commonly occurring sialic acid in nature, but lacks the C-5 amine functionality of neuraminic acid (Figure 1.4).[35] KDN, first discovered in trout egg glycoprotein, is widely expressed in animals and bacteria, however highest expression levels are seen within the lower-vertebrates.[36-37]

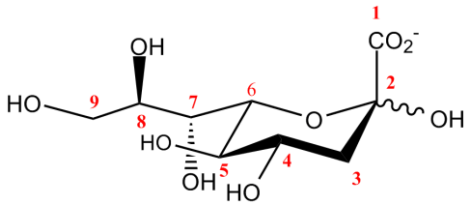


Figure 1.4. Molecular structure of 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid (KDN).

Like NeuNAc, KDN is incorporated into the terminal position of almost all types of vertebral glycoconjugates.[35] Common structural properties shared by these two sialic acids allow a degree of functional overlap; however biological roles unique to KDN arise through loss of the *N*-acetyl group. KDN is known to act as a capping residue of certain vertebral polysialic acid chains, preventing further sialylation of the glycan, and protecting against degradation by bacterial sialidases which specifically recognise terminal NeuNAc residues.[35, 38] KDN is also a component of cell wall polysaccharides in some pathogenic bacteria.[39] Elevated levels of free KDN have been detected in human foetal red blood cells with comparison to maternal red blood cells, and also in ovarian cancer cells, suggesting that KDN may play roles involved in normal development and tumor malignancy.[40]

The biosynthesis of KDN is largely unresolved, and remains a contentious topic. KDN biosynthesis is thought to proceed via a *de-novo* pathway from mannose 6-phosphate (Man 6-P) and PEP, in a process analogous to mammalian NeuNAc biosynthesis.[41] Purified fractions from trout tissue were shown to contain KDN 9-phosphate (KDN 9-P) synthase activity. The authors postulate that this activity was separate to that of NeuNAc 9-PS, which in humans, has secondary KDN 9-P synthase activity.[37, 41] The ability to additionally utilise Man 6-P in catalysis, albeit with far lower efficiency than ManNAc 6-P, gives human NeuNAc 9-PS a unique bi-functional character not yet observed in other mammalian orthologues.[37] To date, no genes encoding a distinct KDN 9-P synthase have been identified, making it difficult to delineate the enzymes of KDN biosynthesis from those involved in NeuNAc biosynthesis.

1.1.5 Legionaminic acid

Legionaminic acid (5,7-diacetamido-3,5,7,9 tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid; Leg) is one of several sialic-acid like sugars often referred to as "bacterial sialic acids".[20] Such compounds are apparently unique to microorganisms and bacteria.[10] Leg is *N*-acetylated at positions C-5 and C-7, and lacks the C-9 hydroxyl of the neuraminic acid core (Figure 1.5). Stereochemically, Leg retains the same absolute configuration as NeuNAc.[10]

Leg is a constituent of the lipopolysaccharides (LPS) of gram-negative pathogenic bacteria including *Legionella pneumophila*, the bacterial pathogen responsible for the acute respiratory condition Legionnaire's disease.[42-43] Leg residues have also been detected within the glycan portion of flagellar glycoproteins from the pathogen C*ampylobacter coli*.[44] Heavy post-translational glycosylation of flagellar proteins is known to be critical for bacterial motility and pathogenicity.[45-47]

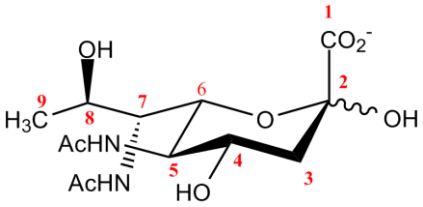


Figure 1.5. Molecular structure of 5,7-diacetamido-3,5,7,9 tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid (legionaminic acid).

A full biosynthetic pathway of Leg has been reported for *Campylobacter jejuni* (Figure 1.6), and was suggested to proceed via unique guanosine-diphosphate (GDP) linked intermediates, differing from the previously described sialic acid biosynthetic pathways which utilise UDP-linked intermediates.[10] The proposed legionaminic acid biosynthetic pathway begins with the glycolytic intermediate fructose 6-phosphate. Seven subsequent reactions yield the metabolite GDP-2,4-diacetamido-2,4,6-trideoxy-*gluco*-pyranose. Simultaneous hydrolysis of the C-1 nucleotide diphosphate (NDP) and epimerisation at C-2 is catalysed by the enzyme GDP-sugar hydrolase/2-epimerase, generating the sugar-substrate 2,4-diacetamido-2,4,6-trideoxy-D-*manno*-pyranose. This compound undergoes condensation with PEP to form legionaminic acid, a reaction catalysed by legionaminic acid synthase (LegS), an enzyme thought be analogous to those that catalyse the PEP condensation reactions of the NeuNAc and KDN biosynthetic pathways.[10] As for NeuNAc and KDN, a final step in the pathway conjugates Leg to CMP, activating it for incorporation into glycoconjugates.[10] It should be noted that in contrast to the proposed Leg biosynthetic pathway described above, Glaze *et al* previously

identified a pathway which instead utilises UDP-linked intermediates.[42] It is likely that more than one NDP is capable of fulfilling this role.

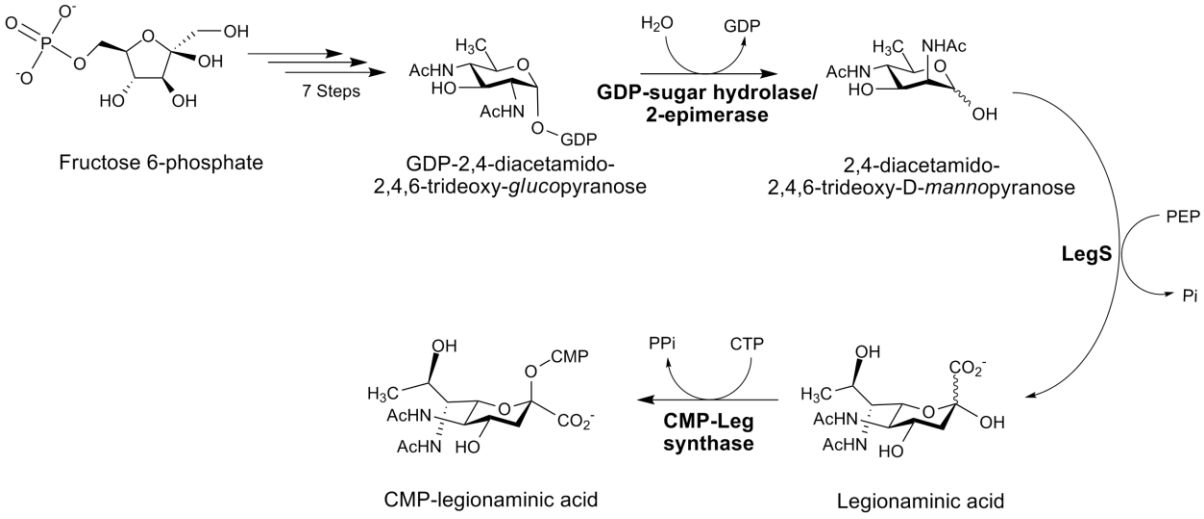


Figure 1.6. The biosynthetic pathway of legionaminic acid as proposed by Schoenhofen *et al*.[10]

1.1.6 Pseudaminic acid

Another common "bacterial sialic acid" is Pseudaminic acid (5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid; Pse). This compound is a diastereomer of Leg, with inversion of stereochemistry at C-5, C-7 and C-8 (Figure 1.7).[9]

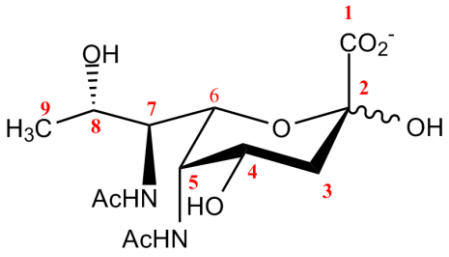


Figure 1.7. Molecular structure of 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid (pseudaminic acid).

Much like Leg, Pse is an important sugar-component of the glycoproteins that make up bacterial flagella.[48] In particular, Pse has been shown to be essential for the proper development of flagella in the gastrointestinal pathogens *C. jejuni*, *C. coli* and *Helicobacter pylori*.[11, 49-51] Functional flagella are required for motility in these organisms, and motility is directly related to their capacity for virulence.[9, 11] Mutant variants of *C. jejuni* and *H. pylori* unable to produce Pse were shown to be aflagellate, highlighting the fundamental role of

glycosylation in flagellar development.[11, 50] This link to bacterial virulence has spurred interest in the pseudaminic acid biosynthesis pathway as a target for novel antibiotics.[9]

The biosynthetic pathway of pseudaminic acid, elucidated first from *H. pylori*, begins with the activated sugar UDP-*N*-acetylglucosamine (UDP-GlcNAc) (Figure 1.8).[7] Three further enzymatic conversions yield UDP-2,4-diacetamido-2,4,6-trideoxyaltrose. A UDP-sugar hydrolase catalyses the cleavage of UDP at C-1, forming 2,4-diacetamido-2,4,6-trideoxyaltrose, which then undergoes a condensation reaction with PEP to form pseudaminic acid and inorganic phosphate.[9] This PEP condensation reaction is catalysed by the enzyme pseudaminic acid synthase (PseS), again thought to be analogous to the PEP utilising enzymes of the previously described sialic acid biosynthesis pathways.[9] CMP-pseudaminic acid synthase conjugates the product to CMP, activating the sugar for transfer to and incorporation into glycoconjugates.[9, 34]

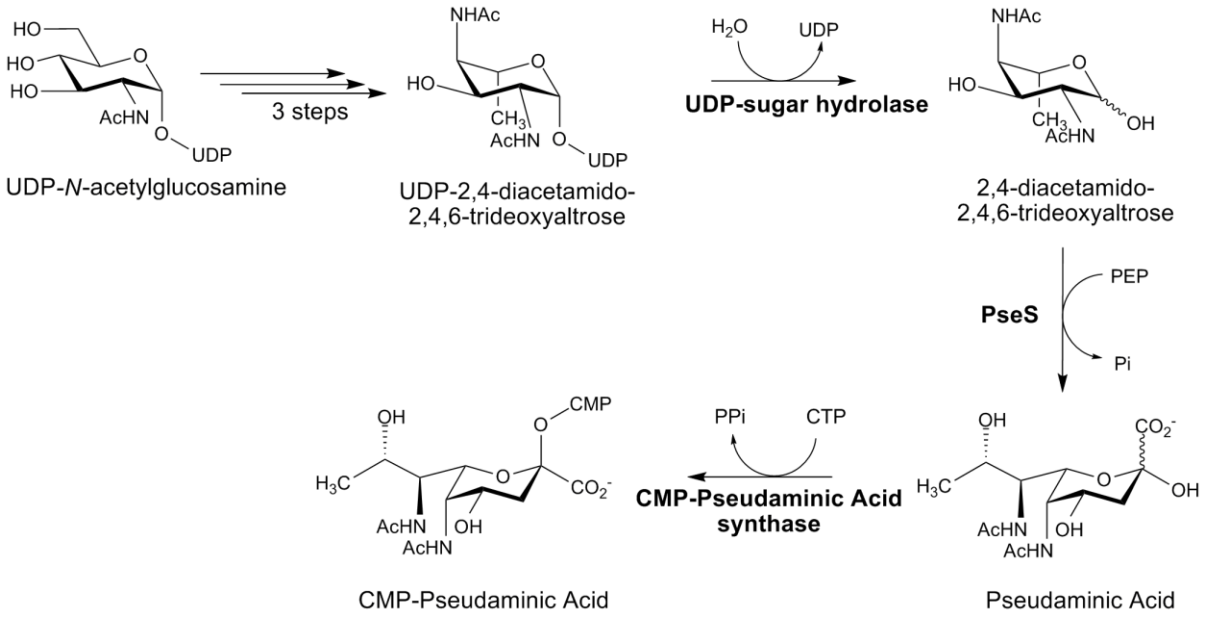


Figure 1.8. The pathway of CMP-pseudaminic acid biosynthesis.

1.2 The sialic acid synthases

The biosynthetic pathways leading to several commonly occurring sialic acids have been described in a number of organisms, as outlined previously. Common to all of these pathways is the central reaction step in which a sugar metabolite undergoes a condensation reaction with PEP to yield the respective sialic acid product. The PEP utilising aldolases which catalyse this reaction can be broadly classified as sialic acid synthases.[1, 52]

1.2.1 N-Acetylneuraminic acid synthase

The most well understood of the sialic acid synthases is *N*-acetylneuraminic acid synthase (NeuNAcS), which has been rigorously characterised from a number of sources.[4, 32, 53-55] In bacterial systems, NeuNAcS catalyses the aldol-like condensation reaction of ManNAc with PEP, forming NeuNAc and inorganic phosphate.[32] In mammals (and a number of other animal phyla), an analogous enzyme catalyses the aldol-like condensation of ManNAc 6-P with PEP, yielding NeuNAc 9-P, and thus this enzyme is referred to as NeuNAc 9-P synthase (NeuNAc 9-PS).[4]

NeuNAcS from the neuroinvasive pathogen *N. meningitidis* is the most highly characterised of the sialic acid synthases.[28, 32, 56] While its identity was unknown at the time, NeuNAcS from *N. meningitidis* (*Nme*NeuNAcS) was first purified from crude cell lysate in 1962, and shown to irreversibly catalyse the condensation reaction of ManNAc and PEP to form NeuNAc.[28] In 1983 a set of genes located within the 17 kilobase "*kps* cluster" of the *E. coli* K1 genome was identified as being necessary for the biosynthesis of the bacteria's polysialic acid capsule.[57] A number of mutations to the genes within this cluster resulted in acapsular mutants and complementation with functionally intact *kps* fragments was able to restore capsulation.[57] One gene located within the *kps* cluster, *NeuB*, was shown to encode NeuNAc synthase activity in *E. coli* K1.[58] Subsequently, the *NeuB* gene was successfully cloned and the gene product purified to homogeneity.[53] NeuNAc synthase activity was then confirmed via a coupled enzymatic assay with CMP-NeuNAc synthetase, which resulted in formation of the end product CMP-NeuNAc.[53] Purified *E. coli* K1 NeuNAcS was shown to be highly specific for the sugar-substrate ManNAc, with a range of alternate hexose and hexosamine sugars not able to be utilised by this enzyme for catalysis.[53]

A gene sequence homologous to *NeuB* from *E. coli* K1 was then identified within the genome of *Streptococcus agalactiae* following a BLAST search of the EMBL/GenBank/DDBJ database.[55] The gene product was purified to homogeneity and shown to facilitate the highly specific condensation of ManNAc and PEP to form NeuNAc. Furthermore, the authors demonstrated that this enzyme adopts a homo-dimeric structure in solution, and has an absolute requirement for divalent metal ions in catalysis, with manganese and magnesium the most activating of those tested.[55] The identification of a *NeuB* homolog within the genome of *N. meningitidis* suggested that a common biosynthetic method may exist between many

species of gram-negative bacteria, and that the NeuNAcS activity originally isolated from *N. meningitidis* cell lysate was likely encoded by this *NeuB* sequence homolog.[59]

1.2.2 Structure of N-acetylneuraminic acid synthase

The molecular structure of *Nme*NeuNAcS (PDB: 1XUZ) was solved in 2005 using X-ray crystallography, and to date remains the only variant of any sialic acid synthase to have its complete structure solved and published.[32] The initial structure was obtained from *Nme*NeuNAcS co-crystallised with Mn^{2+}, PEP and a reduced open chain analogue of ManNAc (rManNAc), all of which are bound within the active site of the protein (Figure 1.9).[32]

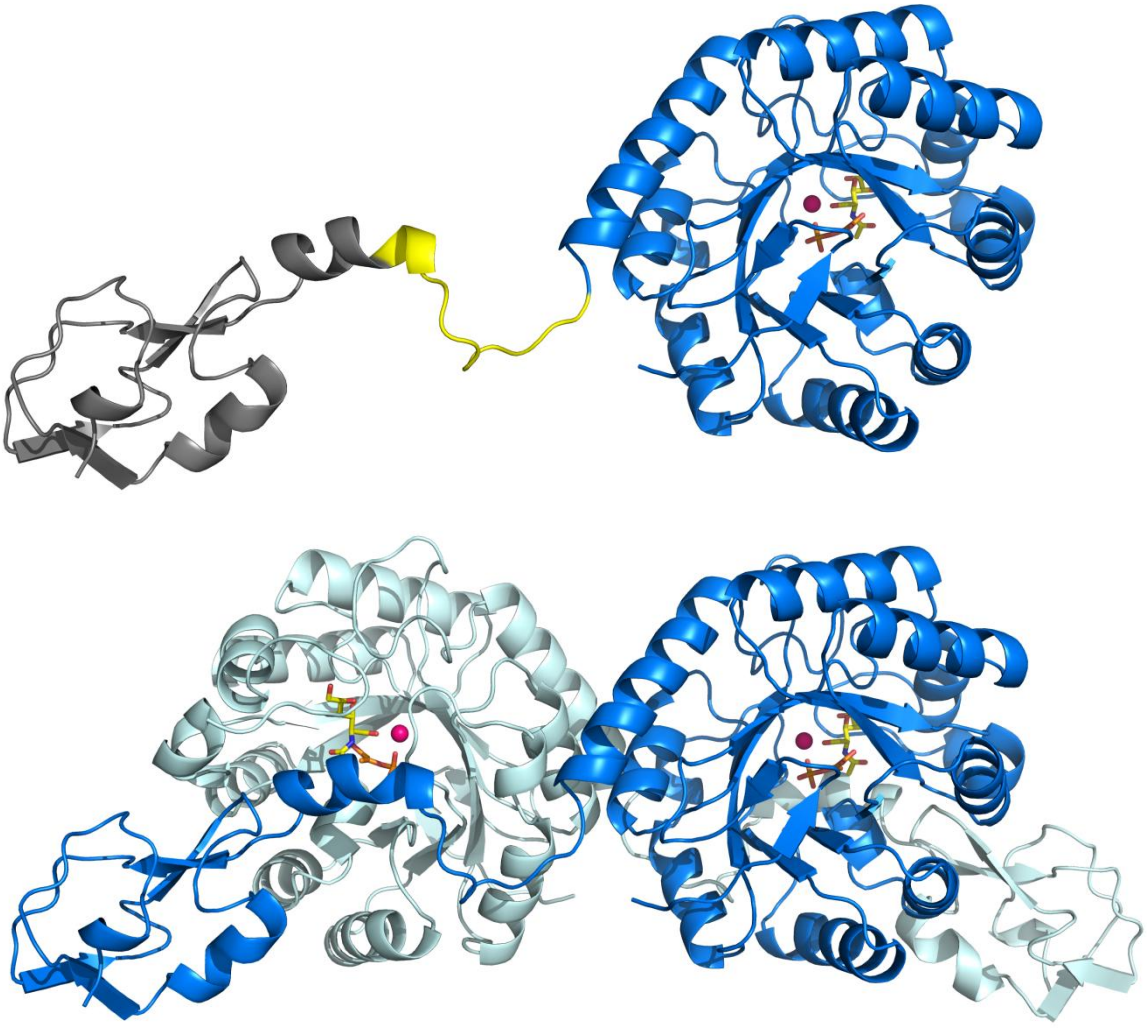


Figure 1.9. The crystal structure of *Nme*NeuNAcS (PDB: 1XUZ). Each monomeric chain (top) consists of an AFPL domain (grey), flexible linker (yellow) and catalytic $(\beta\alpha)_8$ barrel domain (blue). Two identical chains associate to form a domain-swapped homo-dimer arrangement (bottom).

The monomeric chain of *Nme*NeuNAcS was shown to consist of a catalytic N-terminal $(\beta\alpha)_8$ barrel domain (residues 1-273) and a C-terminal antifreeze protein like (AFPL) domain (residues 285-349) joined via a flexible linker region (Figure 1.9). Both of these domains were previously predicted to occur in *Nme*NeuNAcS based on sequence homology with previously solved protein folds, and while the $(\beta\alpha)_8$ barrel fold is common to PEP utilising enzymes (albeit with low sequence identity), the AFPL domain is entirely unique to the sialic acid synthases.[60-62] The AFPL domain is named for its structural homology with the functional type III antifreeze proteins found in arctic fish.[63] These small globular proteins facilitate the inhibition of ice crystal formation, and thus are involved in the freeze-tolerance and freeze-avoidance strategies required for the survival of these organisms in sub-zero environments.[63] Like the type III antifreeze proteins, the AFPL domain of *Nme*NeuNAcS consists of four short β-strands, arranged to generate two-fold symmetry of two identical four-stranded motifs.[32]

Consistent with *in vitro* analysis in solution, *Nme*NeuNAcS was shown to form a homo-dimeric arrangement within the crystals asymmetric unit.[4, 32] Interestingly, the dimer adopts a conformation in which the AFPL domain of one monomer interacts with the $(\beta\alpha)_8$ barrel domain of the other, forming a "domain-swapped" homo-dimer which has since been proven to be critical for catalysis (Figure 1.9).[64] The active site of *Nme*NeuNAcS is housed at the C-terminal face of the $(\beta\alpha)_8$ barrel domain, as is typical for enzymes of the $(\beta\alpha)_8$ barrel fold.[65] The AFPL domain of the opposing monomer effectively caps the active site from solvent, and additionally contributes a key arginine residue (Arg-314) which functions to steer ManNAc into a catalytically favoured orientation within the active site.[64]

The active site of *Nme*NeuNAcS is defined largely by residues from β-strands two and four at the C-terminal of the $(\beta\alpha)_8$ barrel, however the linker region and AFPL domain also contribute residues to the active site (Figure 1.10).[32] *Nme*NeuNAcS is known to be active only in the presence of divalent metal ions, and as such a single manganese ion is present within the active site of the *Nme*NeuNAcS structure.[4, 32] The active site metal ion is coordinated with octahedral geometry by the imidazole moieties of His-215 and His-236, O-2-P of PEP, O-1 of rManNAc and two water molecules.[32] The C-1 hydroxyl of rManNAc coordinates the metal in this structure suggesting that the aldehyde which is present in the native substrate ManNAc may similarly coordinate the metal thus allowing activation of the carbonyl centre for nucleophilic attack by PEP.[32]

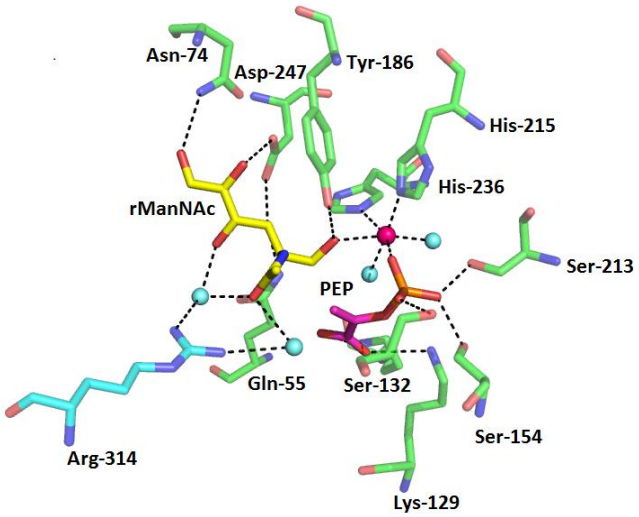


Figure 1.10. Active site depiction of *Nme*NeuNAcS with rManNAc (yellow), PEP (orange) and Mn^{2+} (pink sphere) bound (PDB: 1XUZ). Note the side-chain of Arg-314 (blue) from the opposite monomer extending into the active site where it forms critical interactions with the sugar-substrate via water molecules (blue spheres).

The substrate analogue rManNAc forms interactions with the highly conserved active site residues Asn-74, Asp-247, Gln-55 and Tyr-186 of *Nme*NeuNAcS. Interestingly, Arg-314 from the AFPL domain of the adjacent monomer extends into the active site and forms hydrogen bonding interactions with the *N*-acetyl group of rManNAc via active site water molecules.[32] Deletion of this highly conserved residue renders *Nme*NeuNAcS catalytically inert.[64] PEP also makes a number of interactions within the active site. The phosphate group of PEP forms hydrogen bonds with three conserved serine residues (Ser-132, Ser-154 and Ser-213) and Asn-184, while the PEP carboxylate is shown to form electrostatic interactions with two nearby lysine residues (Lys-129 and Lys-53).[32]

1.2.3 Proposed catalytic mechanism of N-acetylneuraminic acid synthase

NeuNAcS catalyses the aldol-like condensation reaction of PEP with the sugar-substrate ManNAc, to form NeuNAc.[28, 32] Preliminary mechanistic studies on NeuNAcS from *C. jejuni* using deuterated substrates demonstrated that the enzyme catalysed reaction is stereospecific for nucleophilic attack of the *si* face of PEP with the *si* face of the aldehydic (open-chain) form of ManNAc.[54] Subsequent labelling studies with *Nme*NeuNAcS showed that isotopically labelled ^{18}O (bridging C-2 and P atoms) from PEP is released as free phosphate through the course of the catalytic cycle.[32] This observation has important implications for the proposed reaction mechanism. The loss of labelled ^{18}O as free phosphate suggests a mechanism which proceeds via a tetrahedral intermediate.[32] Based on the results of their ^{18}O

labelling studies, Gunawan *et al* proposed a C-O cleavage mechanism (Figure 1.11).[32] Electrons from the enol group of PEP initially attack the C-1 aldehyde of ManNAc, yielding an oxocarbenium ion intermediate. Catalytic water present in the active site then attacks the electrophilic oxocarbenium carbonyl centre, generating a tetrahedral intermediate which spontaneously releases free phosphate (containing the labelled bridging oxygen atom) via C-O bond cleavage. The elimination of phosphate yields open chain NeuNAc which will cyclise to its stable pyranose form in solution. This C-O bond cleavage mechanism is well documented in related PEP utilising enzymes including 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAH7PS) and 2-keto-3-deoxy-D-*manno*-octulosonate-8-phosphate synthase (KDO8PS).[66-67]

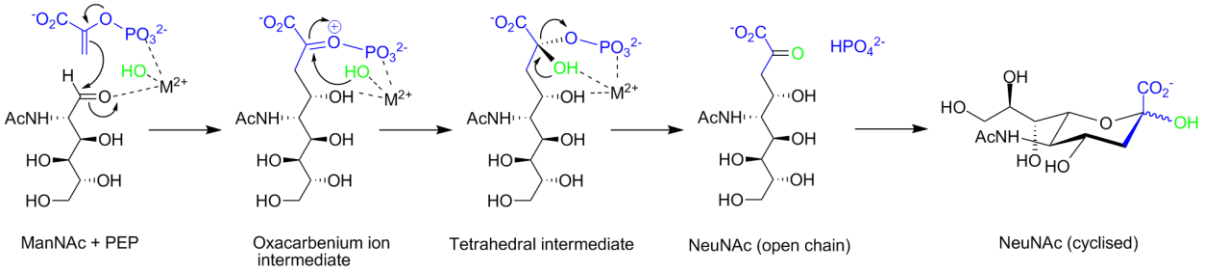


Figure 1.11. The proposed catalytic mechanism of *Nme*NeuNAcS proceeds via oxocarbenium ion and tetrahedral intermediates.

For this reaction mechanism to be plausible, the C-1 carbonyl of ManNAc requires activation to promote nucleophilic attack by PEP in the first step. Similarly the nucleophilicity of the catalytic water must be increased to promote attack of the oxocarbenium ion in the subsequent step. The rManNAc bound crystal structure of *Nme*NeuNAcS (PDB: 1XUZ) shows the active site divalent metal ion occupies a position proximal to the C-1 hydroxyl of the ManNAc analogue.[32] It could be argued that the active site metal may coordinate the aldehyde oxygen of the native substrate, activating the aldehydic carbon to nucleophilic attack by PEP. Several suggestions for the identity of the species responsible for activating water have been made. A glutamate residue (Glu-134) within close proximity of the active site water may act as a base, deprotonating the water and thereby promoting attack on the proximal oxocarbenium ion intermediate.[32] In addition, divalent metal ion coordination of the catalytically relevant water molecule may assist its deprotonation.[32]

The catalytic activity of the sialic acid synthases is widely dependent on the presence of divalent metal ions.[4, 32, 54-55] This metal dependency can be rationalised if the metal is involved in such roles as activation of the ManNAc aldehyde, and the catalytic water.

Crystallisation of *Nme*NeuNAcS in the presence of a stable de-oxy analogue of the supposed tetrahedral intermediate (Figure 1.12) provides strong evidence for a dual role of the divalent metal ion in catalysis.[56]

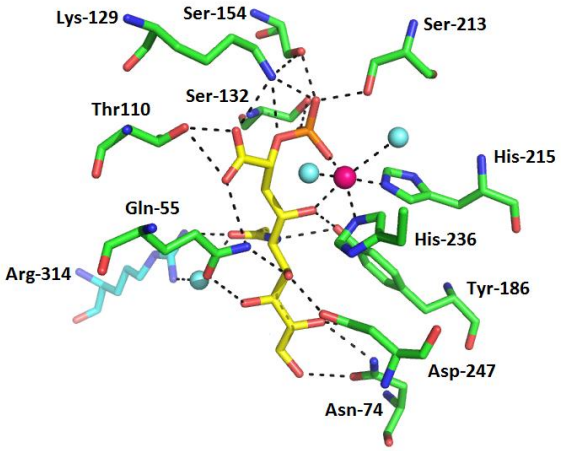


Figure 1.12. *Nme*NeuNAcS active site depiction with putative (*R*)-configuration tetrahedral intermediate analogue (yellow) and Mn^{2+} (pink sphere) bound.

The orientation of the tetrahedral intermediate analogue within the active site strongly implies the role of metal in both activation of the ManNAc aldehyde and delivery of water to the oxocarbenium ion in the following step of catalysis.[56] This inhibitor bound crystal structure also provides rationalisation for the stereochemical preference of the C-2 (*R*) isomer, which was shown to inhibit *Nme*NeuNAcS more potently than its respective (*S*)-diastereomer.[56] In its (*R*) configuration, the C-2 hydrogen projects towards the active site metal ion suggesting that the C-2 hydroxyl of the natural intermediate may coordinate to the metal ion.[56] Preferential binding of the C-2 (*R*) isomer suggests that this is the likely stereochemistry at C-2 of the actual intermediate.[56] The formation of this tetrahedral configuration would require a reaction mechanism that proceeds via the facial attack of catalytic water onto the *si*-face of the oxocarbenium ion intermediate.[56]

Decomposition of the tetrahedral intermediate in the final step of catalysis releases free phosphate and NeuNAc to solution. The driving force behind the collapse of the tetrahedral intermediate is thought to be ionisation of the C-2 hydroxyl group.[32] Promoting deprotonation of the C-2 hydroxyl may be a third role for the coordinated metal ion, or instead the function of a nearby basic residue.[32]

There exists little experimental mechanistic data for any of the other sialic acid synthases, however sequence similarity implies that PEP and sugar-substrate condensation may proceed

via the same mechanism as is now well established for *Nme*NeuNAcS. Glu-25 and Glu-234 are postulated catalytic residues involved in decomposition of the tetrahedral intermediate, these, along with the metal coordinating residues, are almost entirely conserved across phyla, suggesting a mechanistic commonality.[6, 32] Furthermore, the metal dependency of all known sialic acid synthases suggests a common metallo-catalytic cycle. Further elucidation of the precise mechanism of the other sialic acid synthases will be made possible when high resolution crystal structures of these enzymes become available.

The kinetic parameters of the *Nme*NeuNAcS catalysed reaction have been characterised using a number of alternative approaches. Hao *et al* (2005) reported steady-state kinetic parameters for *Nme*NeuNAcS based on a stopped thiobarbituric acid assay (TBA).[4] *Nme*NeuNAcS was shown to have a K_M (ManNAc) of 11.6 mM, suggestive of weak substrate binding. The catalytic turnover number (k_{cat}) of this enzyme was reported as 49.8 catalytic cycles per second.[4] A continuous coupled-assay monitoring the consumption of NADH (a by-product of coupling to NeuNAc lyase and lactate dehydrogenase) at 340 nm was then used to derive the same kinetic parameters.[32] Using this method a K_M (ManNAc) of 9.4 mM and a K_M (PEP) of 0.25 mM were calculated. More recently a continuous UV-visible assay directly monitoring the loss of PEP at 232 nm gave K_M (ManNAc) of 2.9 mM, K_M (PEP) of 0.028 mM and a k_{cat} of 3.1 s^{-1}.[64] These values are lower than those previously reported, and perhaps reflect the increased sensitivity of the continuous assay employed.

1.2.4 N-Acetylneuraminic acid 9-phosphate synthase

The biosynthetic source of NeuNAc in animals was originally thought to be NeuNAc lyase, an enzyme which reversibly catalyses the cleavage of NeuNAc to ManNAc and pyruvate.[26] With equilibrium in favour of the reverse reaction, NeuNAc is formed preferentially. The distinct lack of NeuNAc lyase activity in tissues secreting NeuNAc containing glycoproteins however suggested an alternative route to NeuNAc biosynthesis may operate in animal systems. Indeed, a *de-novo* biosynthetic pathway for NeuNAc was identified in fractions purified from rat liver and bovine submaxillary glands.[29] This pathway consists of three enzymes; a kinase responsible for phosphorylation of ManNAc, an aldolase and a dephosphorylating enzyme. The equilibrium of all three reactions was shown to greatly favour NeuNAc formation. The identified aldolase activity was confirmed in pig submaxillary gland extracts and shown to irreversibly catalyse the formation of NeuNAc 9-P from ManNAc 6-P and PEP, via a proposed condensation reaction mechanism.[27] The aldolase

responsible was tentatively named *N*-acylneuraminic acid synthetase, but is now referred to as *N*-acetylneuraminic acid 9-phosphate synthase (NeuNAc 9-PS), and is known to be the primary facilitator of NeuNAc biosynthesis in animals.[1, 4]

A homology search of the human gene database using the *E. coli NeuB* gene as the search entry identified a homologous sequence (36.1% identity at the amino acid level) encoding a 359 amino acid protein with NeuNAc 9-PS activity.[37] The recombinant gene sequence was shown to restore sialic acid synthase activity in *NeuB*-negative *E.coli* mutants and facilitate NeuNAc formation in insect cells.[37] Interestingly, the same recombinant gene was also shown to cause *in vivo* production of 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid (KDN).[37] In humans, NeuNAc 9-PS has a unique bi-functional character, utilising both ManNAc 6-P and Man 6-P as sugar-substrates for the formation of NeuNAc 9-P and KDN 9-P respectively.[4, 6, 37] It should be noted however that human NeuNAc 9-PS (*Hsa*NeuNAc 9-PS) shows far greater activity in the presence of ManNAc 6-P than Man 6-P and thus is primarily considered a NeuNAc 9-PS.[4]

The steady-state kinetic parameters of various mammalian orthologues of NeuNAc 9-PS have been determined, and the reported parameters vary significantly between studies. The K_M of ManNAc 6-P has been reported as 35 μM, 0.69 mM and 1.04 mM for rat, hog and human NeuNAc 9-PS respectively.[4, 33, 68] The source of such disparity in values may be in the use of relatively insensitive stopped-assay methodologies, or the purity of the enzyme used. The K_M for PEP has been reported as 105 μM for rat NeuNAc 9-PS, and is unreported for any other orthologue.[33] Interestingly, human NeuNAc 9-PS is the only mammalian orthologue to also exhibit activity with Man 6-P, although this secondary activity has been observed in other animals including *Drosophila melanogaster*.[5] The K_M for Man 6-P has been reported as 2.6 mM for *Hsa*NeuNAc 9-PS.[4] It is well accepted that Man 6-P is a far poorer substrate than ManNAc 6-P with respect to the human enzyme (with specificity constants, k_{cat}/K_M, of 4.3 and 21 s^{-1} M^{-1} respectively).[4] It is likely that the absence of the *N*-acetyl group attenuates binding affinity through reduced active site interactions.

NeuNAc 9-PS from a number of sources has been purified and characterised (as outlined above), yet there remains no resolved full length crystal structure of any NeuNAc 9-PS. Considerable sequence homology strongly suggests that *Hsa*NeuNAc 9-PS consists of the same overall folds as elucidated for *Nme*NeuNAcS, a catalytic $(\beta\alpha)_8$ barrel domain and C-

terminal AFPL domain. This enzyme is known to exist as a functional dimer in solution and is expected to form the same domain-swapped homo-dimer arrangement as observed for *Nme*NeuNAcS.[4, 69]

In 2005, the structure of the small C-terminal AFPL domain of *Hsa*NeuNAc 9-PS was solved in solution using NMR spectroscopy (PDB: 1WVO, Figure 1.13).[70] This structure shows overall similarity to the bacterial AFPL domain from *Nme*NeuNAcS with a backbone root-mean-square deviation (RMSD) value of 0.90 Å relative to residues 7-66 from the solved human domain (Figure 1.13).[70] Interestingly the human AFPL domain lacks the conserved residue Arg-314 that enters the active site in the bacterial enzyme. Instead, nearby conserved lysine residues may facilitate the required recognition and stabilisation of the mammalian substrate.[70] Evolutionary analysis of the AFPL domain of human NeuNAc-9-PS and type III antifreeze proteins from both bacteria and vertebrates suggests that the sialic acid synthase domain may be the evolutionary precursor to type III antifreeze proteins in vertebrates.[70] An antifreeze function of the ancestral sialic acid synthase domain may have evolved in response to changes in environmental temperature.[70]

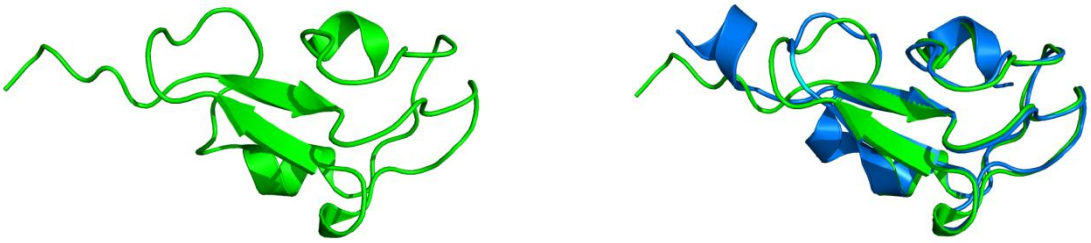


Figure 1.13. Structure of the *Hsa*NeuNAc 9-PS AFPL domain (green, PDB: 1WVO) and structural alignment with the AFPL domain of *Nme*NeuNAcS (blue, PDB: 1XUZ).

1.2.5 Legionaminic acid synthase

The sialic acid synthases responsible for the production of sialic acids other than NeuNAc (i.e. the bacterial sialic acids) are generally lacking in characterisation, with only a handful of studies dedicated to these enzymes. The lack of a commercial supply of the complex sugar-substrates, and the extensive organic synthesis protocol required to arrive at these products has been a major hurdle in the study of the bacterial sialic acid synthases.

The bacterial sialic acid synthase, legionaminic acid synthase (LegS), catalyses the condensation of PEP with 2,4-diacetamido-2,4,6-trideoxy-D-*manno*-pyranose, analogous to the central condensation reaction catalysed by NeuNAcS as previously described.[10] LegS has

been minimally characterised from two bacterial sources (*C. jejuni* and *L. pneumophila*), and was proven to catalyse the aforementioned reaction by NMR, via a metal dependent C-O bond cleavage mechanism similar to that elucidated for *Nme*NeuNAcS.[10, 32, 42] Interestingly, the characterised LegS from *C. jejuni* was also shown to exhibit low levels of NeuNAcS activity suggesting it may be bi-functional in nature.[11]

1.2.6 Pseudaminic acid synthase

Pseudaminic acid synthase (PseS) catalyses the condensation of PEP with 2,4-diacetamido-2,4,6-trideoxyaltrose to form pseudaminic acid (Pse) in certain bacteria.[9, 48] PseS has been characterised from *C. jejuni* and shown to catalyse the formation of Pse by ^{1}H and ^{31}P NMR via the metal dependent C-O bond cleavage mechanism common to the known sialic acid synthases.[48] Kinetic analysis using a coupled enzymatic assay gave a K_M (PEP) value of 6.5 μM, and K_M (2,4-diacetamido-2,4,6-trideoxyaltrose) of 9.5 μM, with a k_{cat} of 0.65 s^{-1}.[48] Unlike LegS, PseS does not exhibit observable secondary NeuNAcS activity.[11]

The *C. jejuni* genome possesses at least three genes encoding sialic acid synthase functionality, *NeuB1*, *NeuB2* and *NeuB3*. The first of these has been characterised as a NeuNAcS, and is 34% identical to *E. coli* NeuNAcS at the amino acid level.[54] *NeuB2* was shown to encode low levels of both LegS and NeuNAcS activity (as described above), and the *NeuB3* gene product has been characterised as a PseS.[42, 48] The fact that this organism contains genes for all three of these related sialic acid synthases makes it an extremely useful system for studying the determinants of sugar-substrate specificity between these enzymes. Understanding the structural basis for selection of one sugar-substrate over another may assist with the design of modified enzymes which could be used to selectively generate sialic acids of novel structure. Furthermore, understanding the ways in which these bacterial enzymes preferentially select and bind their sugar-substrates will assist in the development of potent active-site inhibitors as novel antibiotics.

1.3 Sialic acid synthases as therapeutic targets

Now begins a period of time which will be defined by the distinct threat posed by bacterial infections once treatable with common over-the-counter antibiotics. The rapid emergence of bacterial strains resistant to existing antibacterial compounds, coupled with a dwindling pharmaceutical interest in novel antibiotic development, means the return to a pre-antibiotic era has become a very real possibility.[71]

Antibiotic resistance is not a new phenomenon, nor is it the product of human intervention. In fact, antibiotic resistance genes in bacteria have been present and evolving for more than two billion years.[72] Human interference has however altered the rate at which antibiotic resistance arises. The anthropogenic introduction of massive quantities of antibiotic compounds to the biome via treatment of patients and agricultural livestock provides the strong selection pressures necessary for the development of resistance mechanisms.[71]

A classic example is the now widespread methicillin-resistant *Staphylococcus aureus* (MRSA). MRSA was originally considered the archetypal hospital-superbug, but now exists as a community-acquired pathogen, and additional multi-drug resistant variants are commonplace.[73] Widespread respiratory tract pathogens including *Mycobacterium tuberculosis* and *Streptococcus pneumonia* also exist abundantly in multidrug resistant forms, requiring treatment with drugs previously reserved for last-resort cases.[73] The chief causative agent of bacterial meningitis in children, *N. meningitidis*, is also known to exist as entirely penicillin-resistant forms in some areas and poses an enormous threat, particularly to the parts of Sub-Saharan Africa in which infection is prevalent.[74] So widespread and severe is the emergence of resistance that even last-resort drugs, such as vancomycin and the carbapenems, are ineffective against extremely drug resistant strains. The discovery of a *β*-lactamase in the *enterobacteriaceae* species *Klebsiella pneumonia*, which specifically cleaves the lactam-ring of carbapenems, demonstrates the constant and unrelenting emergence of resistant bacteria, and the ever diminishing supply of drugs with which to treat them.[75]

The development of antibiotic compounds with novel mechanisms of action has never been more important. Circumventing the existing pathways of drug-resistance requires the identification and exploitation of new and unique bacterial targets.

The sialic acid synthases potentially represents a family of novel therapeutic targets for the treatment of bacterial infections otherwise resistant to current therapies. Inhibition of bacterial NeuNAcS activity with high-affinity active site inhibitors would result in reduced NeuNAc expression at the bacterial cell surface, effectively exposing these cells to the host's immune effectors. In this way, sialic acid synthase inhibitors could be considered secondary antibiotics, functioning to destroy the pathogen's molecular disguise, and subsequently facilitating natural clearance by the immune system. Because sialic acid biosynthesis is central to normal human cell function, inhibitors must be entirely bacterio-specific and have no inhibitory effect on NeuNAc 9-PS activity. It is therefore important that we understand the

molecular determinants of substrate specificity between bacterial and mammalian sialic acid synthases to guide the design of inhibitors that will leave the mammalian pathway unperturbed.

The bacterial sialic acid synthases, PseS and LegS, may also provide a unique opportunity for the exploitation of bacterial metabolism by novel drug compounds. Importantly, these enzymes are mostly exclusive to pathogenic gram-negative bacteria and their function is central to motility and virulence.[9-11] Disruption of bacterial cell surface glycoconjugate structures (e.g. LPS and flagella), via inhibition of sialylation, would render these organisms unable to infect their host. Again, the development of inhibitors against these targets first requires an extensive understanding of the enzyme-catalysed reactions and the mode of natural substrate binding.

In addition to this, human NeuNAc 9-PS may hold promise as a novel anti-cancer target. There is significant evidence to suggest that development of malignancy in cancer cells is linked to changes in cellular glycosylation, notably the composition of carbohydrate structures at the cell surface.[76] A number of cancer types including Wilms tumor, colonic carcinoma and small cell lung carcinoma express elevated levels of cell-surface polysialic acid, which may contribute to the overall metastatic potential of these cells.[13] The negative charge conferred by polysialylation of cancer cells may provide the cellular repulsion required to initially detach from the original tumor mass.[13] Once mobile in the blood-stream, polysialic acid rich metastatic cancer cells may interact with sialic acid binding cell-adhesion molecules (e.g. E-selectin) at metastatic sites, facilitating infiltration of the endothelium, and establishment of a nascent metastatic tumor.[13, 76] Furthermore, changes in the sialylation and expression of neuronal cell adhesion molecule promotes metastasis of cancer cells through modulated cell-adhesion processes.[13]

Down-regulation of sialic acid expression at the cell surface through targeted inhibition of the NeuNAc biosynthesis pathway in cancer cells may provide a route to prevention or reduction of tumor metastasis. Further elucidation of the mechanisms linking polysialic acid over-expression and malignancy/metastasis is necessary in order to design effective treatment strategies.

Interestingly, a positive correlation was also detected between levels of free KDN, the alternate product of the *Hsa*NeuNAc 9-PS catalysed reaction, and the malignant state of ovarian adenocarcinomas.[40] Although the link between increased free KDN and tumor

malignancy remains unresolved, KDN levels may find use as a marker of malignant progression of certain ovarian cancers. It has also been suggested that free KDN levels may be an oncofetal antigen for early detection of ovarian cancer and potentially other types of human cancer.[40]

1.4 Research aims

The research undertaken during the course of this study had the general goal of elucidating the molecular basis of substrate specificity of the known sialic acid synthases. By understanding the structural elements which underpin the ability to specifically select for one or more sugar-substrates, we may gain insight into how these enzymes could be exploited as novel therapeutic targets and increase our understanding of how the family of sialic acid synthases have evolved to synthesise the diverse range of compounds we observe in nature. Additionally, modification of the enzyme's structure to accommodate non-natural substrates in catalysis may allow the production of novel complex sialic acids with potentially useful applications.

Before beginning to examine the determinants of sugar-substrate selectivity, I sought to further characterise wild-type human NeuNAc 9-PS, as structural and functional data pertaining to this important enzyme is relatively sparse. This section of research constitutes the second chapter of this thesis and includes: *in vitro* characterisation of the kinetic and thermodynamic parameters of the native reactions catalysed by *Hsa*NeuNAc 9-PS; structural elucidation using homology modelling and mutagenesis; biophysical analysis of the protein's functional arrangement in solution.

In chapter three, I report an attempt to delineate the molecular basis of substrate specificity between *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS, which utilise the substrates ManNAc and ManNAc 6-P respectively. Identification of the structural elements and active site residues which allow the human and bacterial enzymes to select between substrates differing only by phosphorylation at C-6 may provide a basis for the design of bacterio-specific inhibitors which selectively target the bacterial enzyme, whilst leaving the human pathway unperturbed. A bioinformatic approach was taken to identify conserved putative sugar-substrate binding elements from each clade. A number of chimeric proteins combining structural elements from both *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS were then designed and generated in an attempt to interchange the sugar-substrate specificity of these enzymes.

In an extension of the work described in chapter three, the sugar-substrate specificity of two bacterial sialic acid synthases, LegS and PseS, was investigated as detailed in chapter four. The generation of chimeric proteins incorporating a NeuNAcS-like $\beta_2\alpha_2$ loop sequence into these alternative sialic acid synthases was completed in an attempt to confer (or improve) activity with ManNAc, thereby assessing the relative importance of this loop for substrate selectivity. The aim of this work was to identify key sugar-substrate binding motifs which may aid in tracing the evolutionary changes which have occurred allowing these enzymes to utilise alternative sugar-substrates in catalysis and thus become specialised in their biological roles.

Taken together, these three areas of focus aim to provide an increased understanding of the ways in which the sialic acid synthases specifically recognise and bind their appropriate sugar-substrates and provide future direction for the exploitation of these enzymes as either therapeutic targets, or as sources of complex novel sialic acids.

Chapter 2: Structural and functional characterisation of human *N*-acetylneuraminic acid 9-phosphate synthase

2.1 Overview

In mammals, NeuNAc biosynthesis proceeds via the aldol-like condensation of ManNAc 6-P with PEP (Figure 2.1). This central reaction step, catalysed by NeuNAc 9-PS, yields NeuNAc 9-P which is subsequently dephosphorylated to give the final product NeuNAc. This pathway differs from that utilised by bacterial systems, in which unphosphorylated ManNAc undergoes condensation with PEP to give NeuNAc directly. Bacterial NeuNAcS from a number of sources have been characterised in considerable depth however data pertaining to mammalian NeuNAc 9-PS remains relatively sparse.[4, 32, 54-55]

Figure 2.1. The mammalian pathway to NeuNAc biosynthesis.

This chapter details a structural and functional characterisation of human NeuNAc 9-PS (*Hsa*NeuNAc 9-PS). Given that there is currently no full length crystal structure of any NeuNAc 9-PS orthologue available, we report here the generation of a *Hsa*NeuNAc 9-PS homology model based on the structure of *Nme*NeuNAcS, and prediction of the sugar-substrate binding mode, using induced-fit substrate docking. Site-directed mutagenesis was then used to validate the role of predicted active site residues in sugar-substrate binding. Furthermore, a range of biophysical techniques were used to corroborate the overall fold and native quaternary structure of wild-type *Hsa*NeuNAc 9-PS, as predicted by the model.

2.2 Homology modelling of wild-type HsaNeuNAc 9-PS

As there is currently no full length crystal structure of *Hsa*NeuNAc 9-PS, a homology model was generated by Dr. Wanting Jiao, using the PDB coordinates of the solved *Nme*NeuNAcS structure (PDB: 1XUZ) as a template (Figure 2.2). *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS share 29% sequence identity at the amino acid level (Appendix A). The homology model closely resembles the crystal structure of *Nme*NeuNAcS, with root-mean-squared deviation (RMSD) of 0.47 Å, and is predicted to form the same overall fold (Figure 2.3).

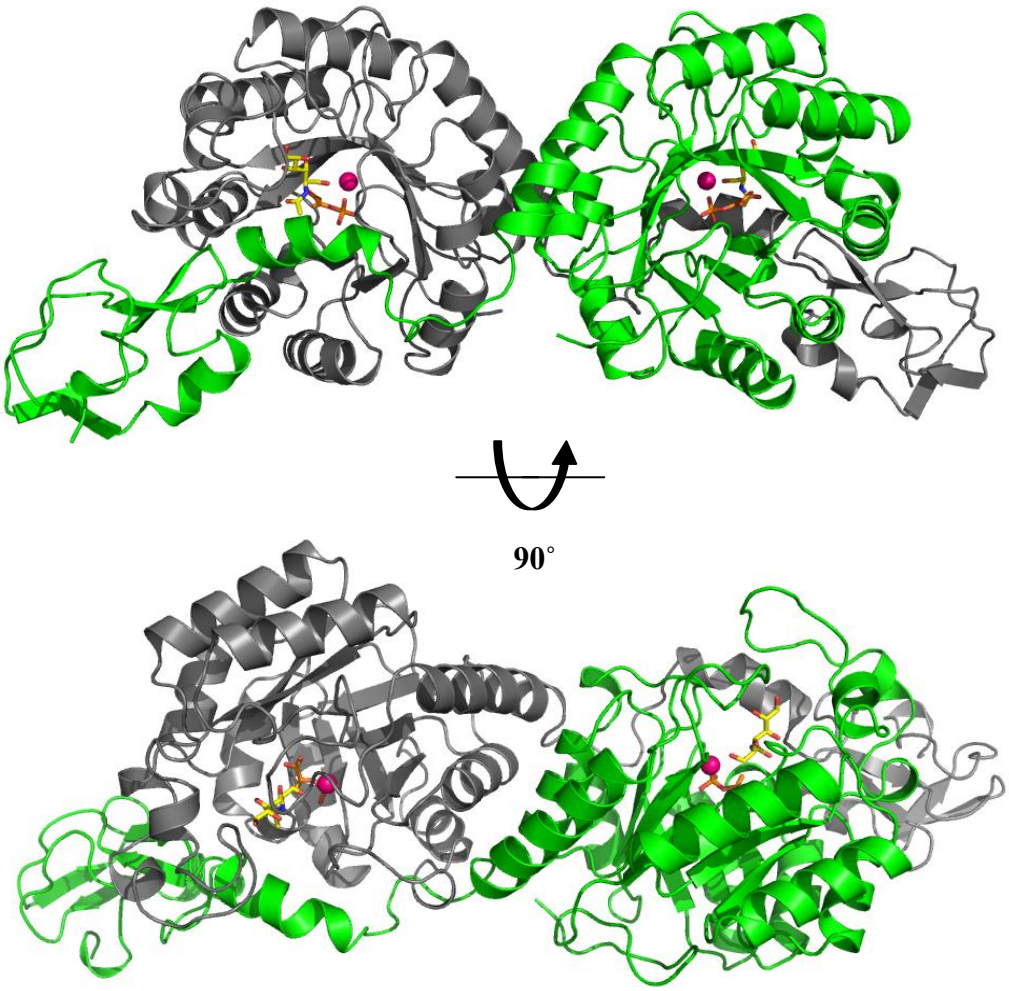


Figure 2.2. Homology model of *Hsa*NeuNAc 9-PS showing domain-swapped homo-dimeric arrangement.

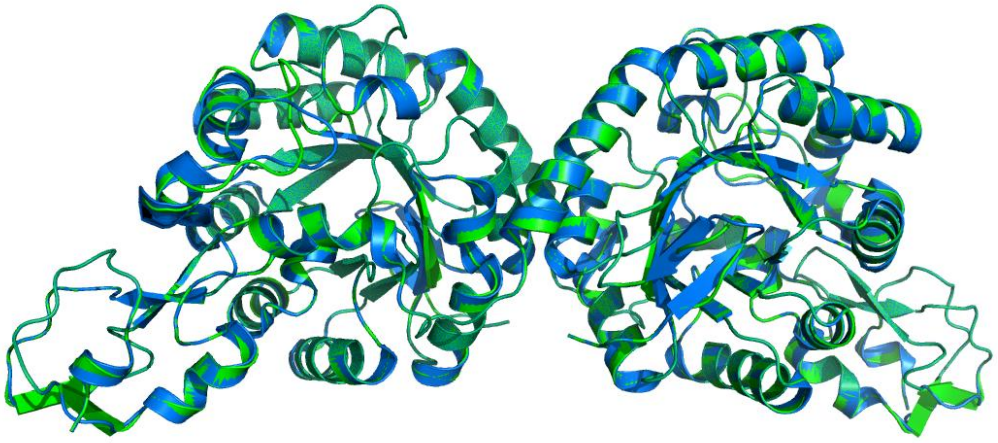


Figure 2.3. Alignment of *Hsa*NeuNAc 9-PS homology model with *Nme*NeuNAcS structure (PDB: 1XUZ).

2.3 Induced-fit substrate docking and rationale for mutagenesis

In order to examine the possible interactions between the substrate ManNAc 6-P and the active site residues of *Hsa*NeuNAc 9-PS, induced-fit docking was conducted by Dr. Jiao, for both chains in the dimer of the model, due to the difference in conformation of the predicted $\beta_2\alpha_2$ loop in both subunits. The induced fit docking in chain A produced 11 output poses, the best scored of which was that with the shortest distance between the reaction centres of PEP and ManNAc 6-P (3.2 Å). Induced-fit docking of ManNAc 6-P into the chain B active site produced four output poses. The difference in the number of output poses between chains is likely due to differences in starting conformations of the active sites in both monomers. The best pose of ManNAc 6-P in chain B has a distance between the two reaction centres of 3.3 Å.

The best poses for both chains predict that PEP and the bound manganese ion form the same interactions as observed in the solved *Nme*NeuNAcS structure, which is not surprising given the conservation of the PEP and metal binding residues across phyla.[6, 32] The active site Mn^{2+} ion is coordinated with octahedral geometry by the imidazole side-chains of His-217 and His-238, the O2P of PEP, O-1 of ManNAc 6-P and two water molecules.[32]

The best ManNAc 6-P binding pose from chain A predicts that the aldehyde oxygen and C-2 acetamide hydrogen bond with the side-chain of Tyr-188 while Gln-54 hydrogen bonds with the C-3 and C-4 hydroxyl groups of ManNAc 6-P (Figure 2.4). The C-4 hydroxyl moiety additionally forms a hydrogen bond with Gln-54 via its amide side-chain. The anionic phosphate group is positioned closely to the positively charged residue Lys-290, with which it is predicted to form an electrostatic interaction. Intriguingly, Lys-290 is inserted into the active site from the AFPL domain of the other monomer, extending across the dimer interface to access the active site of the opposite chain. The phosphate group is also predicted to form hydrogen bonds with His-84 and Thr-80 from the catalytic barrel's $\beta_2\alpha_2$ loop. Overall, this pose of ManNAc 6-P is similar to the binding pose of rManNAc in the crystal structure of *Nme*NeuNAcS (PDB: 1XUZ), with respect to the positions of the *N*-acetyl and aldehyde groups (Figure 2.5).

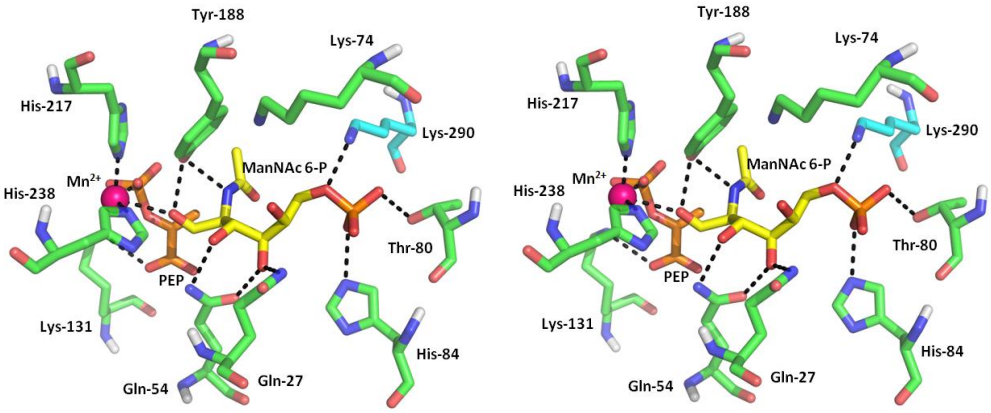


Figure 2.4. Stereo view of the best pose of docked PEP (orange) and ManNAc 6-P (yellow) from chain A. Residues contributed by chain A are shown in green, while residues contributed by chain B are shown in blue.

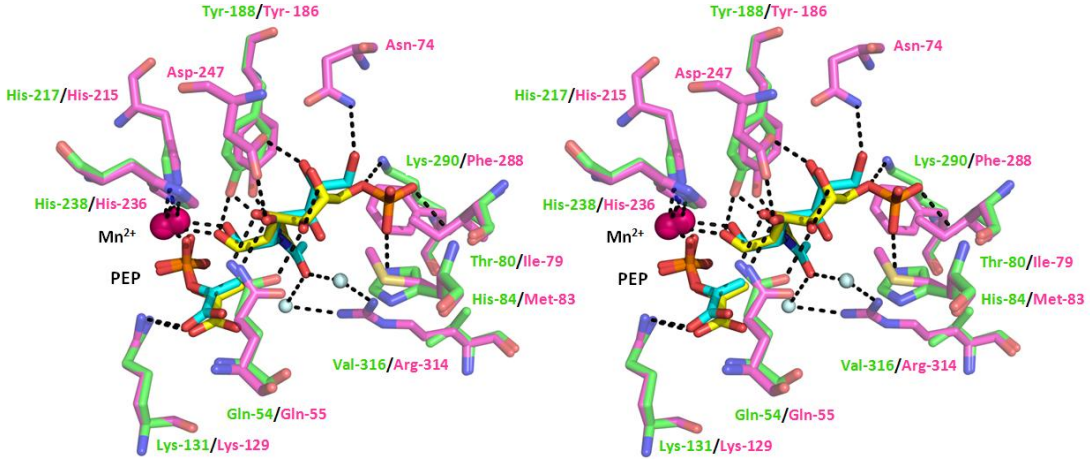


Figure 2.5. Stereo view overlay of Chain A best pose active site residues (green) with *Nme*NeuNAcS (PDB: 1XUZ) active site residues (purple). Mammalian substrates are shown in yellow, while bacterial substrates are coloured blue.

The best output pose for Chain B predicts that the nitrogen of the C-2 acetamide group forms a hydrogen bonding interaction with Tyr-188, the C-3 hydroxyl group forms a hydrogen bond with Gln-54, and the carboxylate side-chain of Asp-247 hydrogen bonds with both the C-2 and C-3 hydroxyl groups of ManNAc 6-P (Figure 2.6). The C-6 phosphate of the mammalian substrate is once again predicted to interact with the positively charged primary amine side-chain of Lys-290, and additionally forms hydrogen bonds with Thr-72.

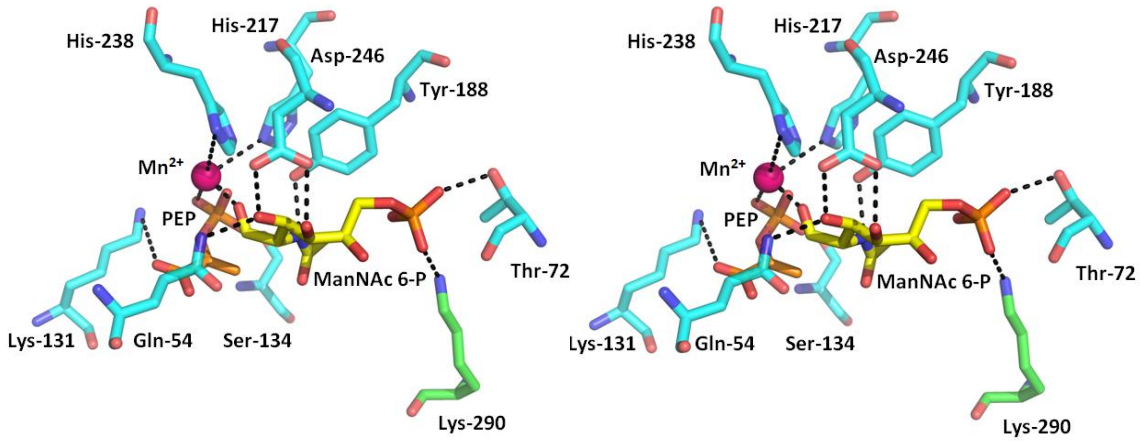


Figure 2.6. Stereo view of the best pose of docked PEP (orange) and ManNAc 6-P (yellow) from chain B. Residues contributed by chain B are shown in blue, while residues contributed by chain A are shown in green. Again, Lys-290 from the AFPL domain of the opposite chain is predicted to form an interaction with the phosphate group of ManNAc 6-P.

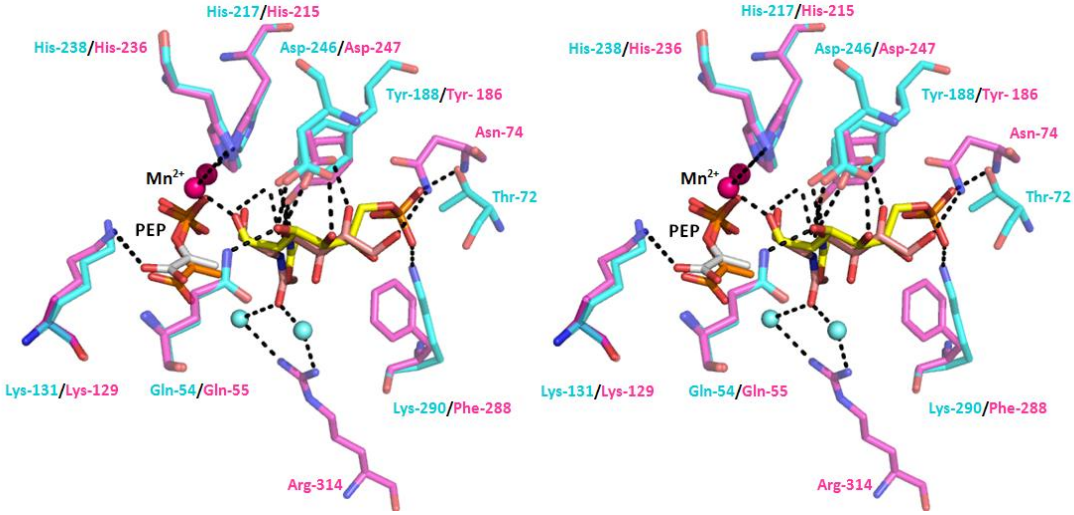


Figure 2.7. Stereo view overlay of Chain B best pose active site residues (blue) with *Nme*NeuNAcS (PDB: 1XUZ) active site residues (purple). ManNAc 6-P is shown in yellow, while ManNAc is shown in pink.

In summary, the ManNAc 6-P binding residues, as predicted by induced-fit docking, are largely the same between the two chains of the homology model. Tyr-188, Gln-55 and Lys-290 are predicted to form bonding interactions in both chains. Additionally, one of two threonine residues (Thr-72 or Thr-80) is predicted to form an interaction with the phosphate group of ManNAc 6-P, and both may contribute some degree of interaction. Of these residues, only Lys-290 and the variable threonine residues are predicted to interact with the C-6 phosphate group which is unique to the mammalian sugar-substrate. Furthermore, these

predicted residues are highly conserved and unique to mammalian NeuNAc 9-PS suggesting involvement in selection for and stabilisation of the phosphorylated sugar-substrate.

In order to verify the role of predicted sugar binding residues, a selection of these were chosen for mutagenesis. The putative phosphate binding residues Lys-290 and Thr-80 were substituted for alanine (which has no capacity for hydrogen bonding) to probe the importance of the respective side-chain functionalities at these positions in sequence. Lys-290 provides a particularly appealing target for mutagenesis, considering it is located on the AFPL domain of the opposite monomer and apparently extends across the dimer interface to form an interaction with the unique phosphate group of the mammalian substrate. This interaction, if real, would have implications for the formation of the domain-swapped homo-dimeric arrangement we predict for *Hsa*NeuNAc 9-PS.

In addition to these two predicted phosphate binding residues, Lys-74 from the $\beta_2\alpha_2$ loop was also chosen for mutagenesis to alanine. While not predicted to form a phosphate binding interaction by substrate docking, this cationic residue occupies a position in sequence which is potentially suggestive of involvement in binding the anionic phosphate group of ManNAc 6-P. Lys-74 aligns closely with Asn-74 from *Nme*NeuNAcS, which is shown to anchor the C-6 hydroxyl of rManNAc in the solved crystal structure. Due to the inherent flexibility of the $\beta_2\alpha_2$ loop, the side-chain of Lys-74 may in-fact be able to adopt an orientation which allows electrostatic interaction with the charged phosphate group of ManNAc 6-P. Such a conformational reorientation may be too large to calculate with the induced-fit docking method used here.

2.4 Cloning, expression and purification of wild-type HsaNeuNAc 9-PS and variants

An *E. coli* codon optimised synthetic gene encoding wild-type *Hsa*NeuNAc 9-PS (GenBank accession No. NP061819) was purchased (Life Technologies) as a Gateway® entry vector pDONR^{TM}221 clone. The target gene was then successfully sub-cloned into the Gateway® expression vector pDEST^{TM}17 which encodes an N-terminal histidine (His) tag as described in section 6.3.3.

The aforementioned variants of *Hsa*NeuNAc 9-PS (K290A, K74A, T80A) were generated by polymerase chain-reaction (PCR) site-directed mutagenesis (Section 6.3.6). The sequences of mutagenic primers used are shown in table 6.5. All mutagenesis was completed using the pDEST^{TM}17 vector bearing the wild-type *Hsa*NeuNAc 9-PS gene as the template, with

PfuUltra DNA polymerase (Agilent). Successful amplification of the mutagenic products was determined by agarose gel electrophoresis, and all mutagenic genes were then sequence verified.

Sequence verified pDEST^{TM}17 vectors bearing wild-type and variant *Hsa*NeuNAc 9-PS genes were then transformed into *E. coli* One Shot TOP10 cells (Invitrogen). Propagated plasmid from a selected transformant was then purified and used for transformation into *E. coli* BL21 (DE3) Star cells and *E.coli* BL21 (DE3) pBB540/pBB542 (Chaperone 3) cells for comparative expression testing. In order to test inducible expression of the target proteins, an induced culture (containing 0.5 mM IPTG) and non-induced culture of each cell line was grown overnight in lysogeny broth (LB). The best yield of soluble protein for wild-type *Hsa*NeuNAc 9-PS and all variants was achieved when expressed in Chaperone 3 cells, and thus this cell line was chosen for up-scaled expression and purification of the desired proteins.

All proteins were expressed and purified using the protocol outlined in section 6.4. Cells over-expressing the target proteins were harvested from 2 L of LB culture by centrifugation, and lysed using sonication. The soluble fraction of the resulting crude lysate was separated by centrifugation, and the His-tagged protein then separated from other soluble cellular protein using immobilized metal-affinity chromatography (IMAC). The N-terminal His-tag was then cleaved by overnight incubation with TEV protease. The cleaved His-tags and TEV protease were then separated from the target protein with another round of IMAC. Size-exclusion chromatography was then used as a final purification step, and the pure fractions pooled, flash frozen and stored at -80˚C. Protein purity was assessed using SDS-PAGE (Figure 2.8A).

Following purification, mass spectrometry was used to confirm the correct molecular weights of wild-type and variant proteins (Appendix B). Circular dichroism (CD) was used to ensure that variant proteins retained wild-type like secondary structure, and were correctly folded in solution (Figure 2.8B).

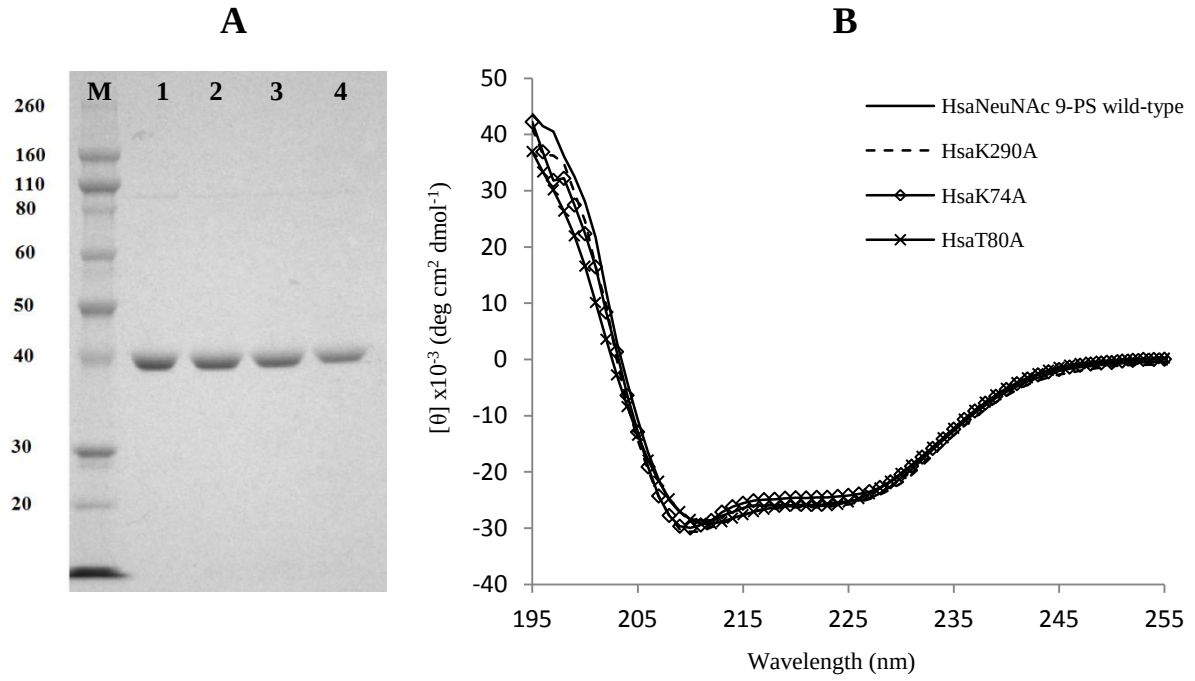


Figure 2.8. A) SDS-PAGE gel showing kDa molecular weight markers (M), purified wild-type *Hsa*NeuNAc 9-PS (1), purified *Hsa*K290A (2), purified *Hsa*K74A (3), and purified *Hsa*T80A (4). **B)** CD spectra of wild-type *Hsa*NeuNAc 9-PS and variants, showing retention of native secondary structure.

2.5 Steady-state kinetics of wild-type HsaNeuNAc 9-PS and variants

The kinetic parameters of purified *Hsa*NeuNAc 9-PS and all mutant variants thereof were determined using a continuous UV-Vis assay monitoring PEP consumption at 232 nm as described in section 6.5.2 (Figure 2.9).

Wild-type *Hsa*NeuNAc 9-PS was shown to have a Michaelis constant (K_M) of 45 ± 3 μM, and k_{cat} of 1.5 ± 0.1 s^{-1} with respect to the sugar-substrate ManNAc 6-P (Table 2.1). The same enzyme was shown to have a significantly lower catalytic efficacy with its alternate substrate Man 6-P. A K_M of 7.3 ± 1 mM and k_{cat} of 0.13 s^{-1} ± 0.01 gives a calculated Man 6-P specificity constant 2000-fold lower than that for ManNAc 6-P (Table 2.1). The absence of the C-2 *N*-acetyl functionality in Man 6-P means substrate-active site interactions are fewer, making it a relatively poor substrate. Furthermore, the measured specificity constant of *Hsa*NeuNAc 9-PS for ManNAc 6-P is more than 300-fold greater than that reported for *Nme*NeuNAcS with ManNAc under similar conditions.[32] Maximal velocities were recorded for *Hsa*NeuNAc 9-PS with PEP concentrations as low as 10 μM (at saturating concentrations of ManNAc 6-P), however an accurate K_M value for PEP could not be determined due to constraints of the assay.

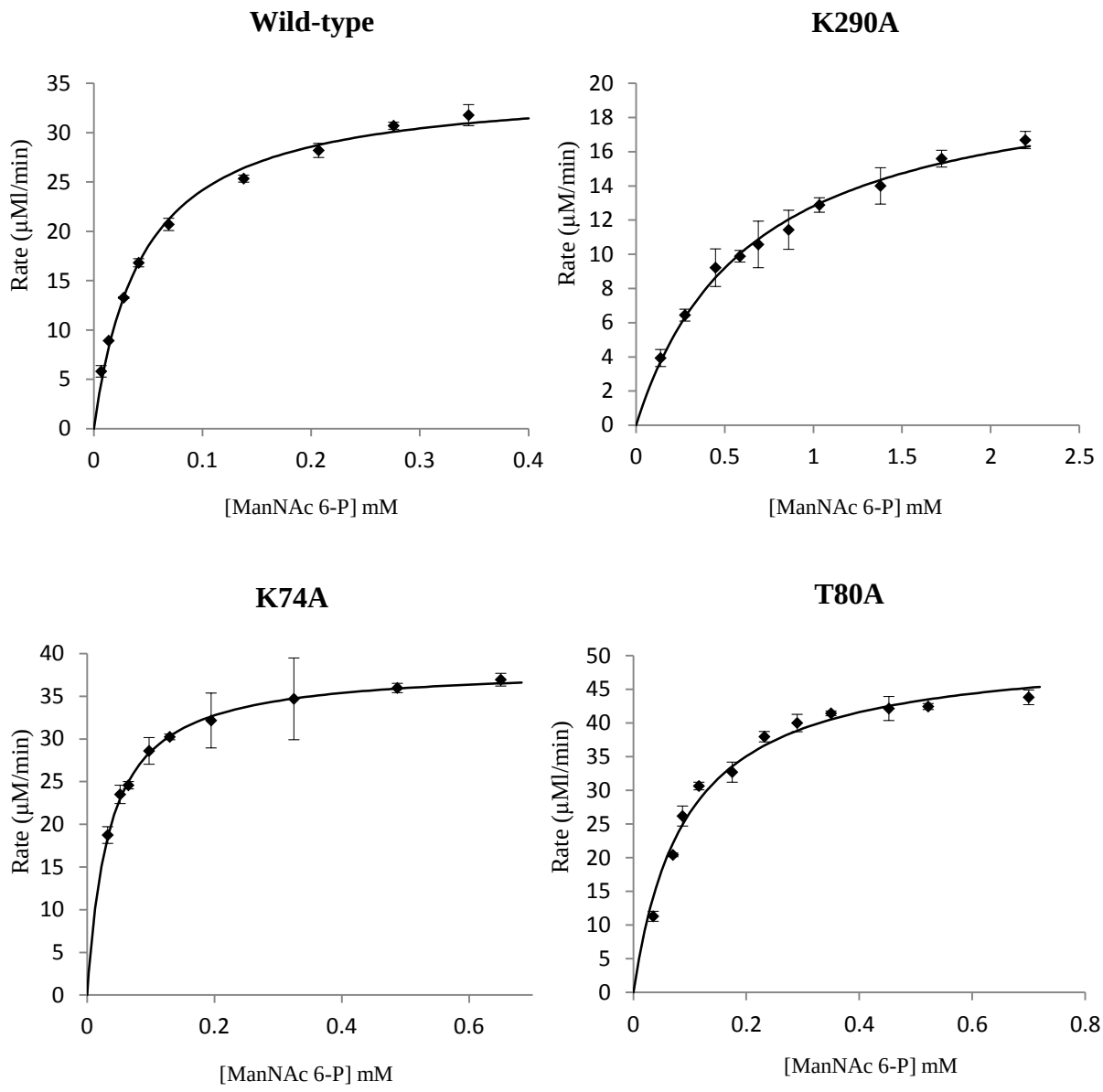


Figure 2.9. Michaelis-Menten plots of wild-type and variant forms of *Hsa*NeuNAc 9-PS at variable concentrations of ManNAc 6-P.

K_M (ManNAc 6-P) of the K290A variant was shown to be considerably increased (approximately 14-fold) over wild-type suggesting attenuated ManNAc 6-P binding, while K_M (PEP) remained less than 10 µM. Similarly, a four-fold reduction in k_{cat} was observed for K290A. Together, an increased K_M (ManNAc 6-P) and reduced k_{cat} results in a >50-fold decrease in the enzymes specificity constant with respect to its native substrate (Table 2.1).

Unlike K290A, the K74A variant displayed near wild-type kinetic parameters (Table 2.1). Loss of the Thr-80 functionality, predicted to hydrogen bond with the C-6 phosphate group of ManNAc 6-P in chain A of the model, results in a two-fold increase in K_M (ManNAc 6-P), and a resultant decrease in the enzymes specificity constant for this substrate (Table 2.1).

Table 2.1. Kinetic parameters of wild-type *Hsa*NeuNAc 9-PS and mutant variants

Kinetic Parameter	Wild-type	K290A	K74A	T80A
K_M **(ManNAc 6-P) μM**	45 ± 3	650 ± 50	35 ± 1	92 ± 10
k_{cat} **(ManNAc 6-P)** s^{-1}	1.5 ± 0.1	0.4 ± 0.1	1.6 ± 0.1	1.4 ± 0.1
k_{cat}/K_M **(ManNAc 6-P)** $s^{-1}M^{-1}$	34000 ± 4000	590 ± 60	46000 ± 1600	14700 ± 2500
K_M **(Man 6-P) mM**	7.3 ± 1			
k_{cat} **(Man 6-P)** s^{-1}	0.13 ± 0.01			
k_{cat}/K_M **(Man 6-P)** $s^{-1}M^{-1}$	18 ± 2			

2.6 PEP binding studies of wild-type HsaNeuNAc 9-PS and variants

Given that K_M (PEP) is apparently very low (< 10 μM) and was not accurately quantifiable using the standard continuous assay, isothermal titration calorimetry (ITC) was employed to determine the dissociation constants (K_d) and binding enthalpies of PEP for wild-type *Hsa*NeuNAc 9-PS and variants thereof (Table 2.2). It was observed that PEP binds with a 1:1 molar ratio, as expected given each monomer of *Hsa*NeuNAc 9-PS contains a single PEP binding site (Figure 2.10A). PEP binding is exothermic for *Hsa*NeuNAc 9-PS, with an overall enthalpic change (ΔH) of -2380 ± 13 cal/mol suggesting favourable hydrogen bond formation. Furthermore, PEP binding is associated with a favourable entropic change (-TΔS) of -5250 cal/mol, which may be indicative of displacement of solvent from the binding site. Interestingly, *Nme*NeuNAcS also exhibits entropically favoured PEP binding, but lacks the favourable enthalpic component shown here for *Hsa*NeuNAc 9-PS (Section 3.9). This additional favourable enthalpy change may account for the lower dissociation constant (and therefore higher binding affinity) of PEP observed for *Hsa*NeuNAc 9-PS over *Nme*NeuNAcS.

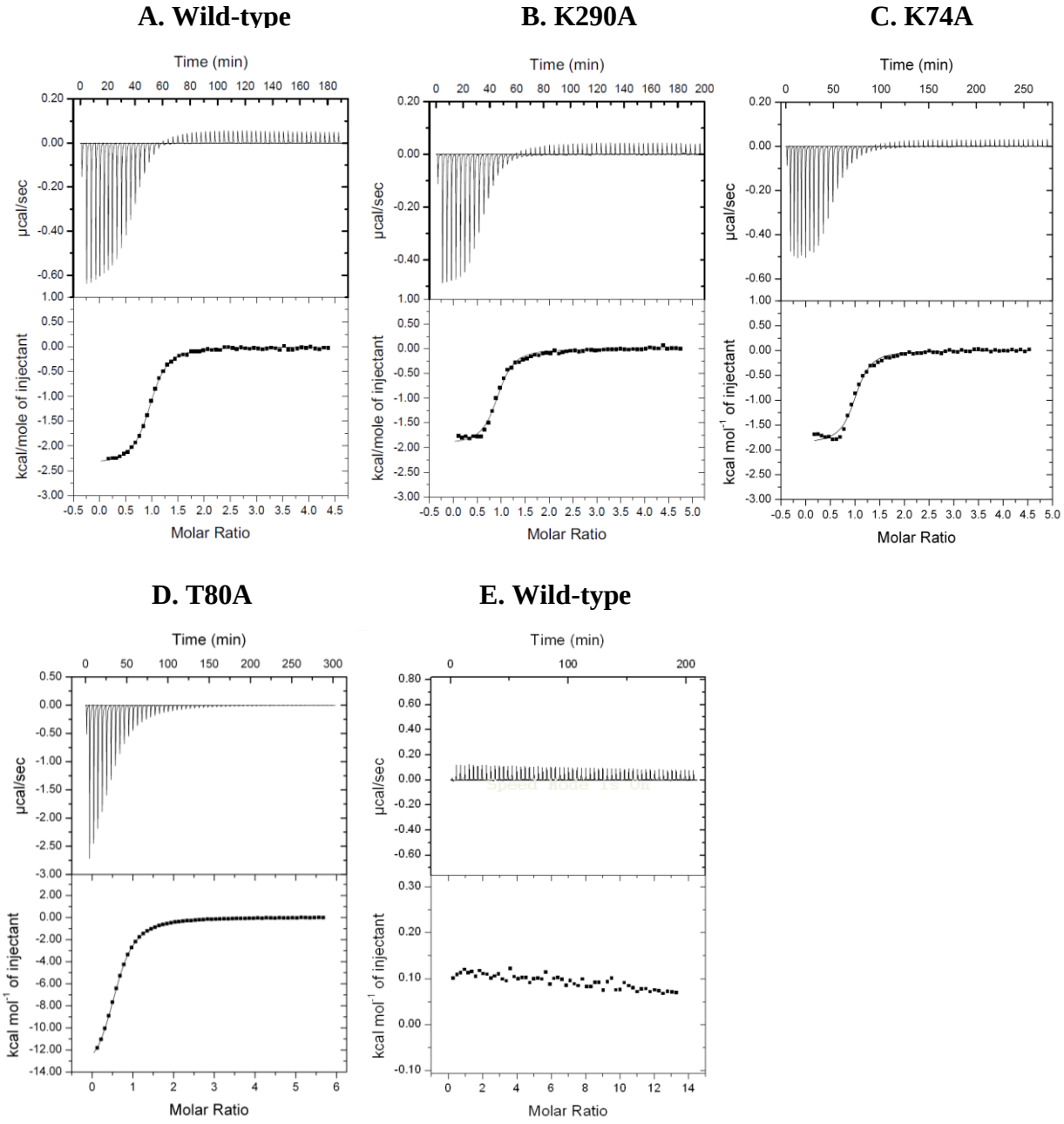


Figure 2.10. (A-D) Raw ITC thermograms and model-fitted data for 2mM PEP titrations into 90 µM wild-type *Hsa*NeuNAc 9-PS and variants. **(E)** Raw ITC thermogram and integrated data for 5mM ManNAc 6-P titration into 90 µM wild-type *Hsa*NeuNAc 9-PS.

The K_d (PEP) for wild-type *Hsa*NeuNAc 9-PS was measured as 2.6 ± 0.1 µM, consistent with a K_M of less than 10 µM. All mutant variants exhibited K_d values consistent with the wild-type enzyme, suggesting introduction of these specific mutations has not significantly perturbed PEP binding (Table 2.2). It is interesting to note that while *Hsa*T80A retains a near wild-type dissociation constant for PEP, the thermodynamic parameters of PEP binding are apparently altered for this variant (Table 2.2). It should be noted however that the full sigmoidal binding curve was not observed for *Hsa*T80A and so the calculated

thermodynamic parameters should be viewed with scepticism (Figure 2.10D). Titration of 5 mM ManNAc 6-P into *Hsa*NeuNAc 9-PS, in the absence of PEP, did not elicit an observable binding event (Figure 2.10E) suggesting PEP is required for the subsequent binding of ManNAc.

Table 2.2. Calculated thermodynamic parameters of PEP binding to wild-type *Hsa*NeuNAc 9-PS and variants.

	Wild-type *Hsa*NeuNAc 9-PS	K290A	K74A	T80A
K_d (PEP) µM	2.6 ± 0.1	3.2 ± 0.3	2.4 ± 0.3	9.2 ± 0.1
ΔH cal/mol	-2380 ± 13	-2010 ± 30	-1880 ± 30	-14870 ± 55
ΔS cal/mol/deg	17.6 ± 2	18.4 ± 2	19.4 ± 2	-26.8 ± 3

2.7 Thermal stability

The thermal stability of wild-type *Hsa*NeuNAc 9-PS and variants was assessed using differential scanning fluorimetry (DSF) in the presence (and absence) of metal and ligands (Figure 2.11). Wild-type *Hsa*NeuNAc 9-PS was shown to have a melt temperature of 43 ˚C, the same as that reported previously for wild-type *Nme*NeuNAcS.[77]

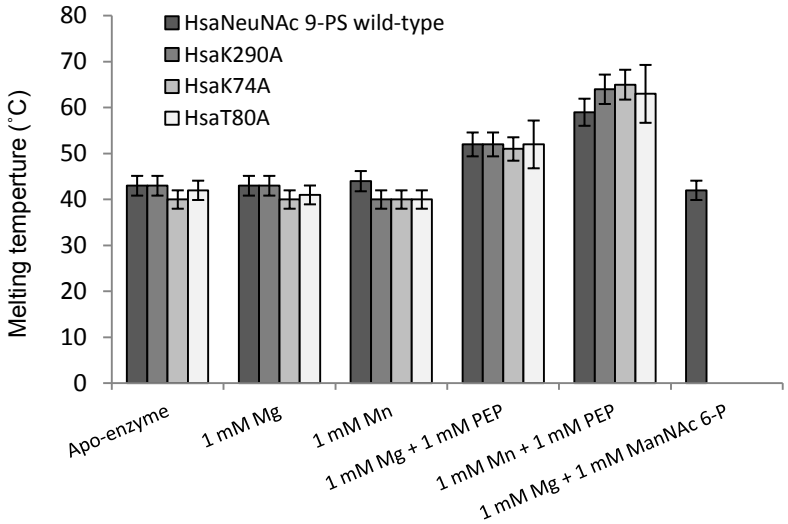


Figure 2.11. The effect of ligands on the thermal stability of wild-type *Hsa*NeuNAc 9-PS and variants.

All variant proteins exhibited similar melting temperatures to wild-type *Hsa*NeuNAc 9-PS under each set of conditions. While no significant thermal shift was observed upon the addition of metal alone, the addition of 1 mM metal and 1 mM PEP results in a substantial

increase in the thermal stability of both wild-type and variants. This result is consistent with a structural rearrangement upon PEP binding, possibly related to pre-organisation of the sugar binding site for ManNAc 6-P binding. Interestingly, the combination of Mn^{2+} and PEP confers almost twice the degree of thermal stabilisation as Mg^{2+} and PEP. No thermal stabilisation was observed when the protein was supplemented with metal and ManNAc 6-P. This is likely indicative of the inability of ManNAc 6-P to bind in the absence of PEP, corroborating the lack of ManNAc 6-P binding observed by ITC.

2.8 Structural studies

A number of in-solution techniques were used to validate the overall fold and quaternary structure predicted for *Hsa*NeuNAc 9-PS by homology modelling. It has been previously reported that the wild-type enzyme exists predominantly as a native dimer in solution.[69] The quaternary structure of all variants of *Hsa*NeuNAc 9-PS was also assessed to ensure that any loss of functionality observed was not the result of disrupted oligomerisation.

The native quaternary structure of wild-type *Hsa*NeuNAc 9-PS and variants was assessed in solution by analytical gel filtration on a Superdex 200 10/300 column. Protein standards of known molecular weights were first used to generate a semi-logarithmic calibration curve of elution volume/void volume (V_e/V_o) versus molecular weight (Figure 2.12).

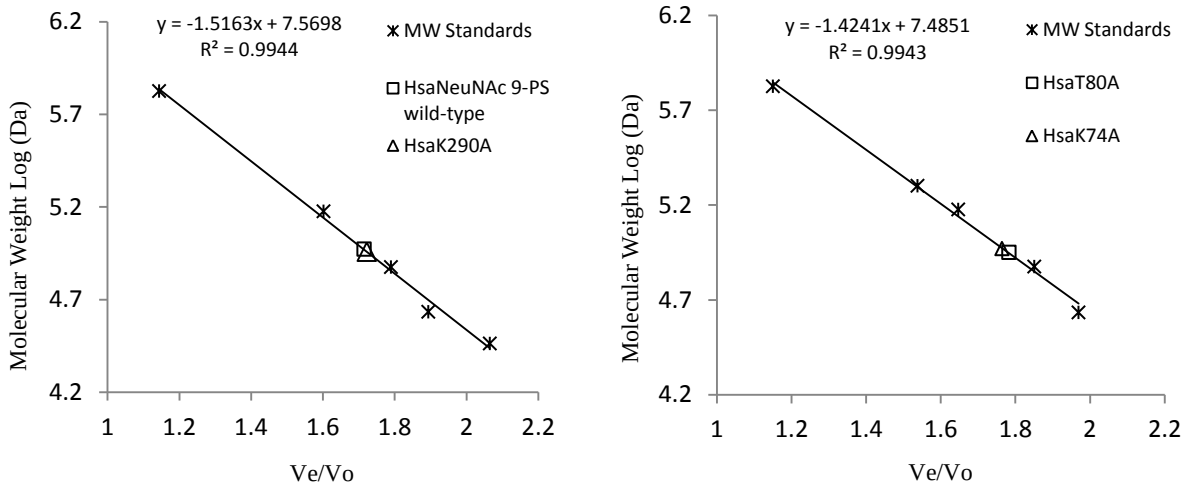


Figure 2.12. Semi-logarithmic plots showing elution volume of target proteins relative to molecular weight standards. Standards used for calibration include at least five of the following: thyroglobulin (669 kDa), beta amylase (200 kDa), alcohol dehydrogenase (150 kDa), ovalbumin (43 kDa), conalbumin (75 kDa) and carbonic anhydrase (29 kDa).

The elution volume of His-tagged wild-type *Hsa*NeuNAc-9-PS was measured and converted to molecular mass by substitution into the equation describing the linear calibration curve. The His-tagged wild-type enzyme elutes as a single species at 13.5 mL, with a calculated molecular weight of ~94 kDa (Table 2.3). This value is consistent with that of a dimeric species in solution (theoretical dimeric mass: 87.7 kDa), and corroborates a previous study suggesting *Hsa*NeuNAc 9-PS exists primarily as a functional dimer.[69] All mutant variants of *Hsa*NeuNAc 9-PS were also shown to exist as dimeric species by analytical gel filtration under the experimental conditions used, confirming that introduction of respective sugar binding site mutations has not perturbed the enzymes quaternary structure.

Table 2.3. Experimental parameters determined by analytical gel filtration

	Elution volume (mL)	Elution vol/void vol	Calculated molecular mass (Da)	Theoretical dimeric mass (Da)	Error (%)
***Hsa*NeuNAc 9-PS wild-type (His-tagged)**	13.5	1.71	93800	87732	6.4
***Hsa*K290A (His-tagged)**	13.6	1.72	90900	87619	3.6
***Hsa*K74A (His-tagged)**	13.3	1.76	93300	87619	6.1
***Hsa*T80A**	13.4	1.78	89100	80931	9.2

Small angle X-ray scattering (SAXS) was then used to probe the solution structure of *Hsa*NeuNAc 9-PS, and visualise any conformational change associated with PEP binding. Given the thermal stabilisation induced by PEP, and the observation that ManNAc 6-P does not bind in the absence of PEP, it was hypothesised that PEP binding induces a conformational active site rearrangement which primes the enzyme for ManNAc 6-P binding. A PEP induced conformational change has been described previously for *Nme*NeuNAcS, with a marked change in the SAXS profile observed upon PEP binding.[77]

SAXS data was collected for *Hsa*NeuNAc 9-PS at approximately 10 mg/mL, in the presence and absence of 1mM $MnCl_2$ and PEP. Experimental scattering data was then compared to the theoretical scattering profile of the *Hsa*NeuNAc 9-PS homology model using CRYSOL.[78] Figure 2.13 shows the experimental scattering data for both unliganded and liganded *Hsa*NeuNAc 9-PS overlaid with the theoretical homology model scattering. Samples were

devoid of an increase in intensity at low q (indicative of aggregation) with Guinier plots linear for $qR_g < 1.3$.

Both apo and liganded protein data sets are in good agreement with the predicted homology model scattering, with chi-square values of 1.1 and 1.2 respectively. There is however no significant change in the scattering profiles of apo and PEP bound *Hsa*NeuNAc 9-PS.

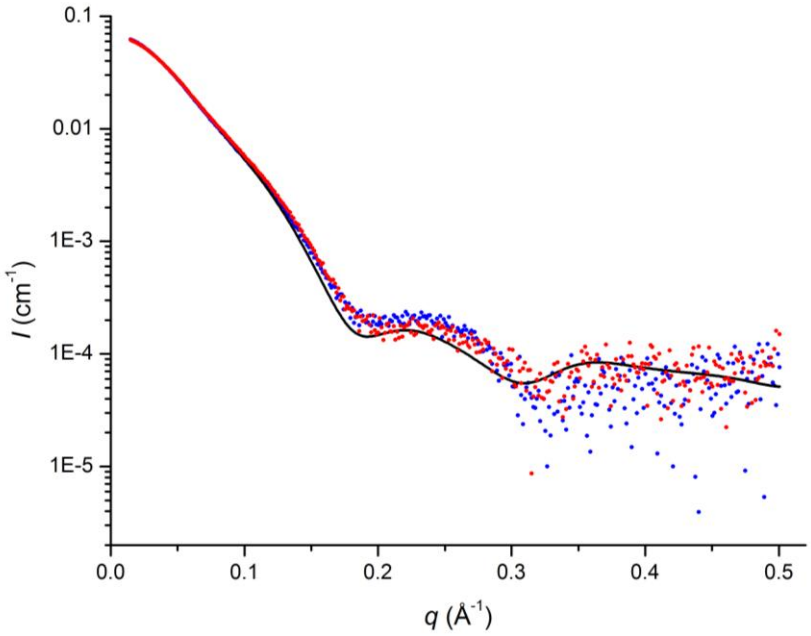


Figure 2.13. Small angle X-ray scattering (SAXS) profiles of apo (blue) and PEP bound (red) *Hsa*NeuNAc 9-PS, fit to the predicted homology model scattering profile (black line).

Using the structural parameters derived from SAXS experiments (Table 2.4), we are able to estimate overall volume and molecular mass of the protein in solution. The porod volume (V_P) estimate obtained from scattering data is 106300 ± 11000 Å^{-3} and 115600 ± 12000 Å^{-3} for liganded and unliganded wild-type *Hsa*NeuNAc 9-PS respectively. This consistent with the theoretical volumes calculated from sequence. Using the online server MoW, this volume can be used to calculate the approximate molecular mass of the protein in solution.[79] The calculated molecular mass of both liganded and unliganded *Hsa*NeuNAc 9-PS is approximately twice the theoretical monomeric mass, confirming dimerisation in solution. Close agreement of the experimental scattering profiles with the predicted model scattering strongly implies *Hsa*NeuNAc 9-PS forms the domain-swapped homo-dimeric arrangement that is predicted by the homology model, and has been observed previously for

*Nme*NeuNAcS.[32] Such domain-swapping would allow the previously identified phosphate binding residue Lys-290 to extend into the active site from the AFPL domain of the opposite chain as predicted.

Table 2.4. Experimental parameters obtained by small angle X-ray scattering (SAXS).

Structural parameters	*Hsa*NeuNAc 9-PS + $MnCl_2$ + PEP	*Hsa*NeuNAc 9-PS + $MnCl_2$
I(0) (cm^{-1}) [from P(r)]	0.07 ± 0.01	0.05 ± 0.01
R_g (Å) [from P(r)]	34.3 ± 1	34.8 ± 1
I(0) (cm^{-1}) (from Guinier)	0.07 ± 0.01	0.05 ± 0.01
D_{max} (Å)	118 ± 6	119 ± 6
R_g (Å) (from Guinier)	33.8 ± 0.3	34.3 ± 0.3
Porod volume estimate ($Å^3$)	106300 ± 11000	115600 ± 12000
Dry dimeric volume calculated from sequence ($Å^3$)	97,540	97,540
Calculated monomeric molecular mass from sequence (Da)	40495	40495
Molecular mass from Porod volume (Da) (from MoW)	88260	95940

2.9 Crystallisation trials

Extensive crystallisation trials were conducted for wild-type *Hsa*NeuNAc 9-PS and variants. These trials included screening commercial crystallisation conditions (PACT *premier*, Clear Strategy I & II and JCSG Plus) at variable protein concentrations in the presence and absence of metal and PEP. While no crystals of sufficient diffraction quality were obtained from these trials, a number of positive leads were identified:

1) Poor quality plate-like crystals formed after approximately seven days when fresh *Hsa*NeuNAc 9-PS (6.95 mg/mL) was crystallised by hanging drop vapour diffusion method in the presence of 0.2 M sodium bromide, 0.1 M bis-tris propane (pH 7.5), 20 % w/v PEG 3350, 1 mM PEP and 1 mM $MgCl_2$ (Figure 2.14A). Similar crystals were obtained with alternate sodium salts (sodium nitrate, sodium formate, sodium acetate trihydrate). Diluting the entire condition to 85% of the original concentration yielded significantly improved crystals, but they continued to lack three-dimensionality (Figure 2.14B). Hanging drop vapour diffusion was completed in 24-well plates with a total reservoir volume of 500 µL and drop volume of 2 µL.

2) Needle-like crystals were observed after several days when fresh *Hsa*NeuNAc 9-PS (6.95 mg/mL) was crystallised by sitting drop vapour diffusion method with conditions from Clear Strategy I (well E1) and PACT *premier* (well C5) from 96-well plates (Figures 2.14C, 2.14D respectively). The latter diffracted to 3.4 Å but were not cryo-protected and did not survive 90% attenuation. 96-well plates were screened using a TTP Labtech Mosquito screening robot, with final drop volumes of 400 nL. Optimisation of the aforementioned crystal leads was not completed, given time constraints.

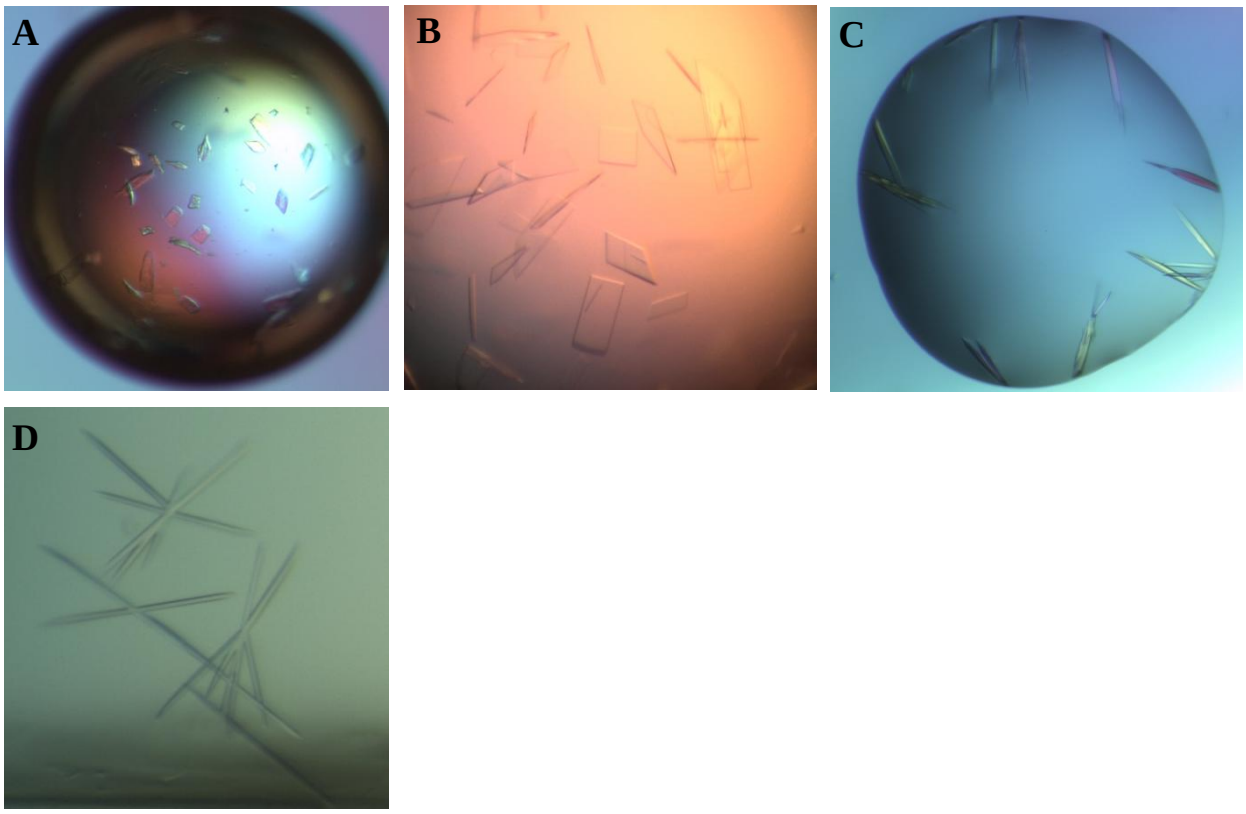


Figure 2.14. Initial crystallisation leads for wild-type *Hsa*NeuNAc 9-PS. **A)** Plate-like crystals grown in 0.2 M sodium bromide, 0.1 M bis-tris propane (pH 7.5), 20 % w/v PEG 3350, 1 mM PEP and 1 mM $MgCl_2$ (2 µL drop). **B)** Improved plate-like crystals grown in 85% final concentration of the condition used in A (2 µL drop). **C)** Needles grown in conditions from Clear Strategy I - E1 (400 nL drop). **D)** Needles grown in conditions from PACT *premier* - C5 (400 nL drop).

2.10 Discussion

Unlike bacterial NeuNAcS, which utilise the neutral amino sugar ManNAc in catalysis, mammalian NeuNAc biosynthesis proceeds via condensation of PEP with the phosphorylated substrate ManNAc 6-P. By utilising a charged, phosphorylated sugar-substrate, the mammalian enzyme gains increased substrate specificity over its bacterial counterpart. This is evident in the massive reduction of the human enzymes K_M (ManNAc 6-P), and subsequent

rise in specificity constant for ManNAc 6-P, relative to that of *Nme*NeuNAcS for ManNAc.[32, 64] The high affinity interactions afforded by the additional phosphate group allows the mammalian enzyme to be more specific for its substrate, which is particularly important in a cellular environment rich in alternate hexose and hexosamine sugars, and contributes to the overall precision of sialic acid metabolism in mammalian systems.

Lys-290, from the AFPL domain of the opposite monomer, in part contributes to the selectivity for this unique mammalian substrate, by forming a high affinity interaction with the C-6 phosphate group. Given the abundance of cationic residues present on the proximal $\beta_2\alpha_2$ loop, it is interesting to discover that electrostatic stabilisation of the sugar-substrate phosphate group is achieved partially through interaction with this lysine residue from the adjacent monomer. Insertion into the active site from a location distant in sequence is reminiscent of Arg-314, a conserved bacterial NeuNAcS AFPL domain residue known to be critical for catalysis in *Nme*NeuNAcS. The insertion of Lys-290 into the active site from the AFPL domain of the adjacent monomer provides rationalisation for the formation of the functional domain swapped homo-dimeric arrangement. It is likely that monomeric variants of *Hsa*NeuNAc 9-PS would have reduced or abolished catalytic ability through lack of this 'borrowed' active site residue.

The binding mode of ManNAc 6-P, shown here to rely on a Lys/Thr pair for phosphate binding shares some commonality with related PEP utilising aldolases. 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAH7PS) for example utilises the phosphorylated sugar-substrate erythrose 4-phosphate, and a conserved Arg/Thr or Arg/Ser pair, also located on the $\beta_2\alpha_2$ loop of the enzymes catalytic $(\beta\alpha)_8$ barrel domain, fulfils the phosphate binding role.[67, 80-81] The key difference however is that the conserved pair of phosphate binding residues in DAH7PS are adjacent in sequence. In *Hsa*NeuNAc 9-PS, Thr-80 is indeed adjacent in sequence to a lysine residue, but this lysine is not predicted to interact with the phosphate group. Instead, Lys-290, distant in sequence to the functional threonine, is apparently recruited for ManNAc 6-P binding.

Substrate binding to *Hsa*NeuNAc 9-PS appears to be sequential, with no ManNAc 6-P binding event observable in the absence of PEP. It is possible that PEP binding drives pre-organisation of the active site, which promotes subsequent sugar-substrate binding, as previously observed for *Nme*NeuNAcS.[77] The thermal stabilisation conferred by PEP provides some evidence for an induced conformational rearrangement or structural

rigidification, however the lack of change between liganded and unliganded SAXS profiles suggests that any such rearrangement does not significantly affect the average global structure of the enzyme. The favourable entropic change associated with PEP binding (also observed for *Nme*NeuNAcS) is likely to involve displacement of water from the active site upon binding. The entropic gain achieved by solvent exclusion would potentially compensate any entropically disfavoured conformational rearrangement associated with PEP binding. Interestingly, PEP binding to the human enzyme is also driven by an enthalpic component not observed for *Nme*NeuNAcS.

To conclude, we have identified a number of key residues likely to be involved in the selection and stabilisation of the phosphorylated sugar-substrate of *Hsa*NeuNAc 9-PS. Further elucidation of the specific substrate binding interactions will require access to a high resolution crystal structure.

Chapter 3: Determinants of the sugar-substrate specificity of bacterial and mammalian sialic acid synthases

3.1 Overview

As previously described, the pathways of NeuNAc biosynthesis differ between mammalian and bacterial systems.[1, 4] In bacteria, NeuNAcS catalyses the aldol-like condensation of ManNAc and PEP to form NeuNAc directly.[32] The mammalian PEP condensing enzyme, NeuNAc 9-PS, is however unable to utilise ManNAc in catalysis. Instead, ManNAc must first undergo phosphorylation to ManNAc 6-P by ManNAc kinase.[29] This product is the substrate for NeuNAc 9-PS, which catalyses condensation with PEP to give NeuNAc 9-P. Dephosphorylation in a subsequent step forms the functional product NeuNAc.

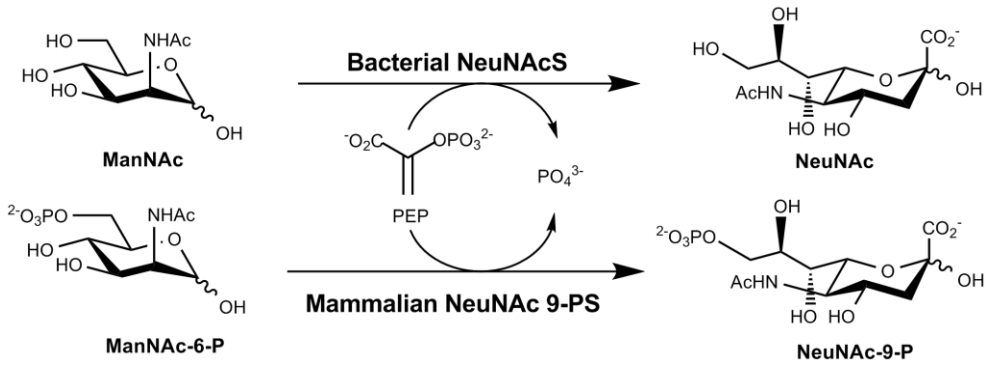


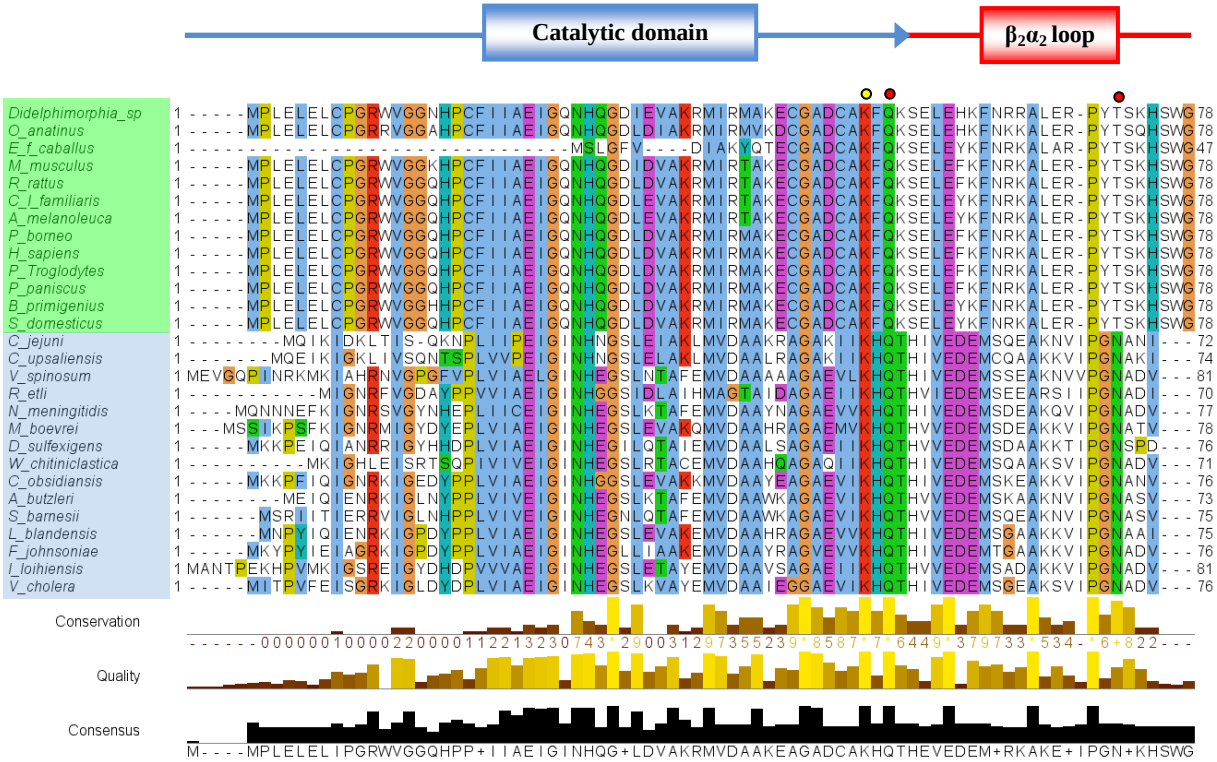
Figure 3.1. The condensation reactions catalysed by bacterial and mammalian sialic acid synthases.

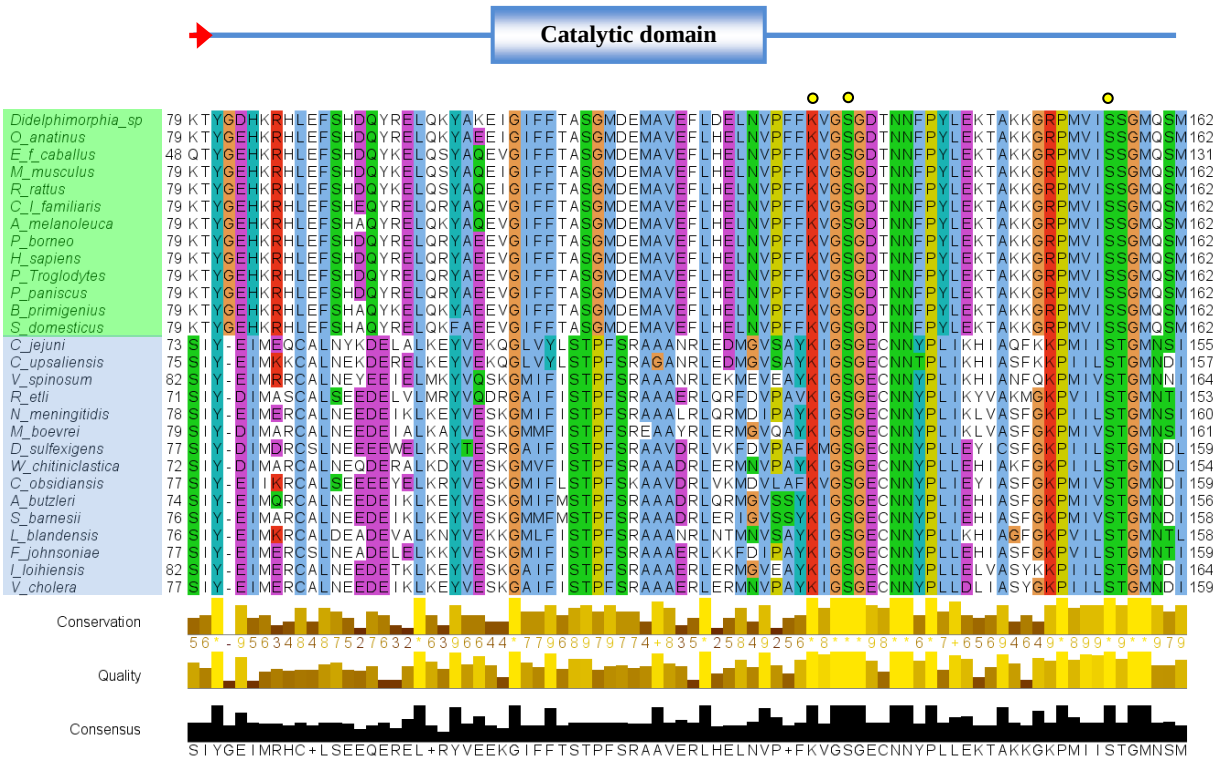
Most remarkably, there is no sugar-substrate ambiguity between the bacterial and mammalian forms, with all characterised bacterial NeuNAcS exclusively utilising ManNAc as sugar-substrate, and likewise the mammalian NeuNAc 9-PS utilising only ManNAc 6-P in catalysis (Figure 3.1).[33, 68] It is apparent that the sugar-substrate specificity of the bacterial and mammalian sialic acid synthases is dependent entirely upon the presence or absence of the C-6 phosphate group. It is important to reiterate here that human NeuNAc 9-PS is also able to utilise mannose 6-phosphate as an alternate substrate.[6, 37] This compound lacks the C-2 *N*-acetyl functionality of ManNAc, but retains the C-6 phosphate group which is apparently an absolute requirement for sugar-substrates of the mammalian enzymes. Understanding the structural determinants which underpin this differential sugar-substrate selectivity is pertinent to the development of novel antibiotic compounds which exclusively target the bacterial enzyme, whilst leaving the human enzyme unperturbed.

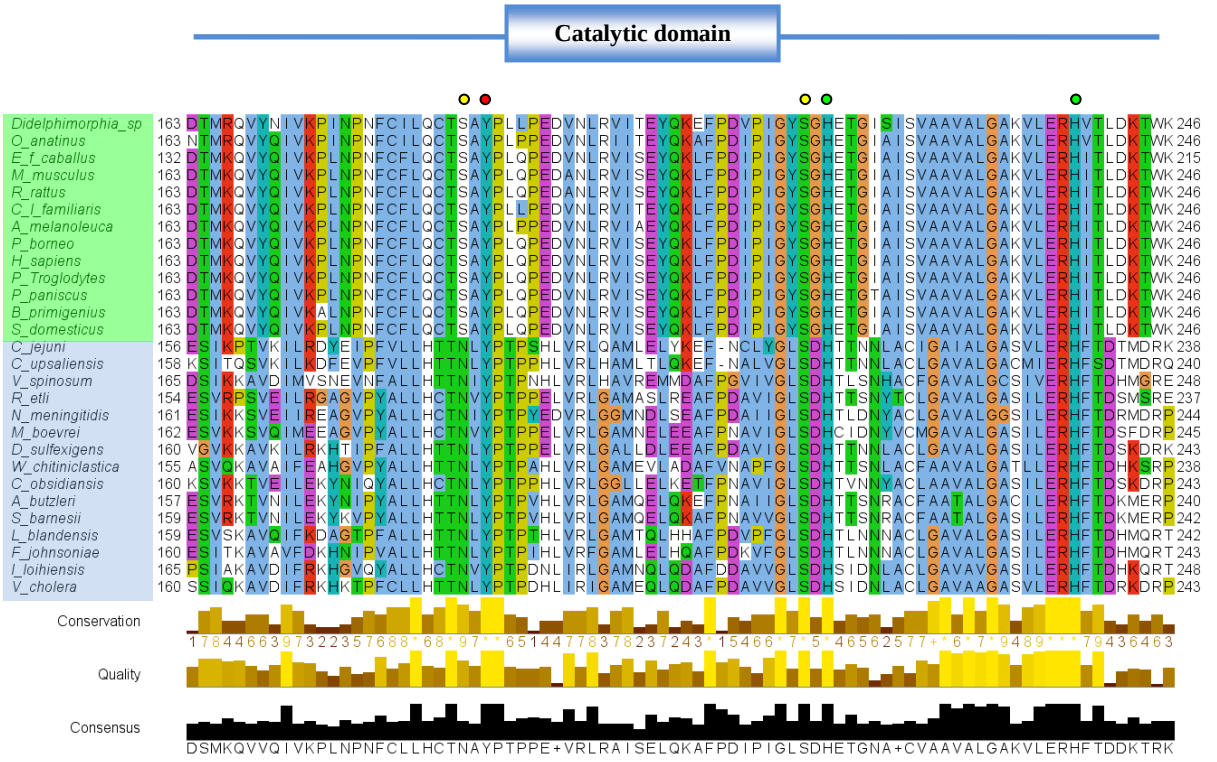
In this chapter, I report the details of a bioinformatic analysis of bacterial and mammalian sequences conducted in order to identify elements which are highly conserved within bacterial and mammalian clades, but poorly conserved between them. Such elements may correspond to structural motifs involved in recognising and stabilising the phosphorylated or unphosphorylated forms of ManNAc. Following the elucidation of two regions of distinct sequence variability between bacterial and mammalian sialic acid synthases, a number of chimeric proteins incorporating structural elements from both *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS were generated and characterised.

3.2 Sequence alignment of NeuNAcS and NeuNAc 9-PS

A selection of 15 bacterial NeuNAcS and 13 mammalian NeuNAc 9-PS sequences were obtained from the KEGG database [82] and input for multiple alignment using Clustal Omega (Figure 3.2).[83-84] While all but one of the mammalian sequences used in this alignment were 359 residues in length, the bacterial sequences vary between 343-352 residues. The additional length of the mammalian enzymes typically arises from an extended C-terminus, with an extra 6-8 residues from the AFPL domain. Furthermore, the $\beta_2\alpha_2$ loop of the bacterial sequences (as defined by the crystal structure of *Nme*NeuNAcS, PDB: 1XUZ) is typically two residues shorter than the corresponding mammalian loop sequence.




Catalytic domain
Didelphimorphia_sp
O_anatinus
E_f_caballus
M_musculus
R_rattus
C_l_familiaris
A_melanoleuca
P_borneo
H_sapiens
P_Troglodytes
P_paniscus
B_primigenius
S_domesticus
C_jejuni
C_upsaliensis
V_spinosum
R_etli
N_meningitidis
M_boevrei
D_sulfexigens
W_chitiniclastica
C_obsidiansis
A_butzleri
S_barnesii
L_blandensis
F_johnsoniae
I_loihiensis
V_cholera
Conservation
Quality
Consensus
SIYGEIMRHC+LSEEQEREL+RYVEEKGIFFTSTPFSRAAVERLHELNVP+FKVGSGECNNYPLLEKTAKKGKPMIISTGMNSM


Catalytic domain
Didelphimorphia_sp
O_anatinus
E_f_caballus
M_musculus
R_rattus
C_l_familiaris
A_melanoleuca
P_borneo
H_sapiens
P_Troglodytes
P_paniscus
B_primigenius
S_domesticus
C_jejuni
C_upsaliensis
V_spinosum
R_etli
N_meningitidis
M_boevrei
D_sulfexigens
W_chitiniclastica
C_obsidiansis
A_butzleri
S_barnesii
L_blandensis
F_johnsoniae
I_loihiensis
V_cholera
Conservation
Quality
Consensus
DSMKQVVQIVKPLNPNFCLLHCTNAYPTPPE+VRLRAISELQKAFPDIPIGLSDHETGNA+CVAAVALGAKVLERHFTDDKTRK

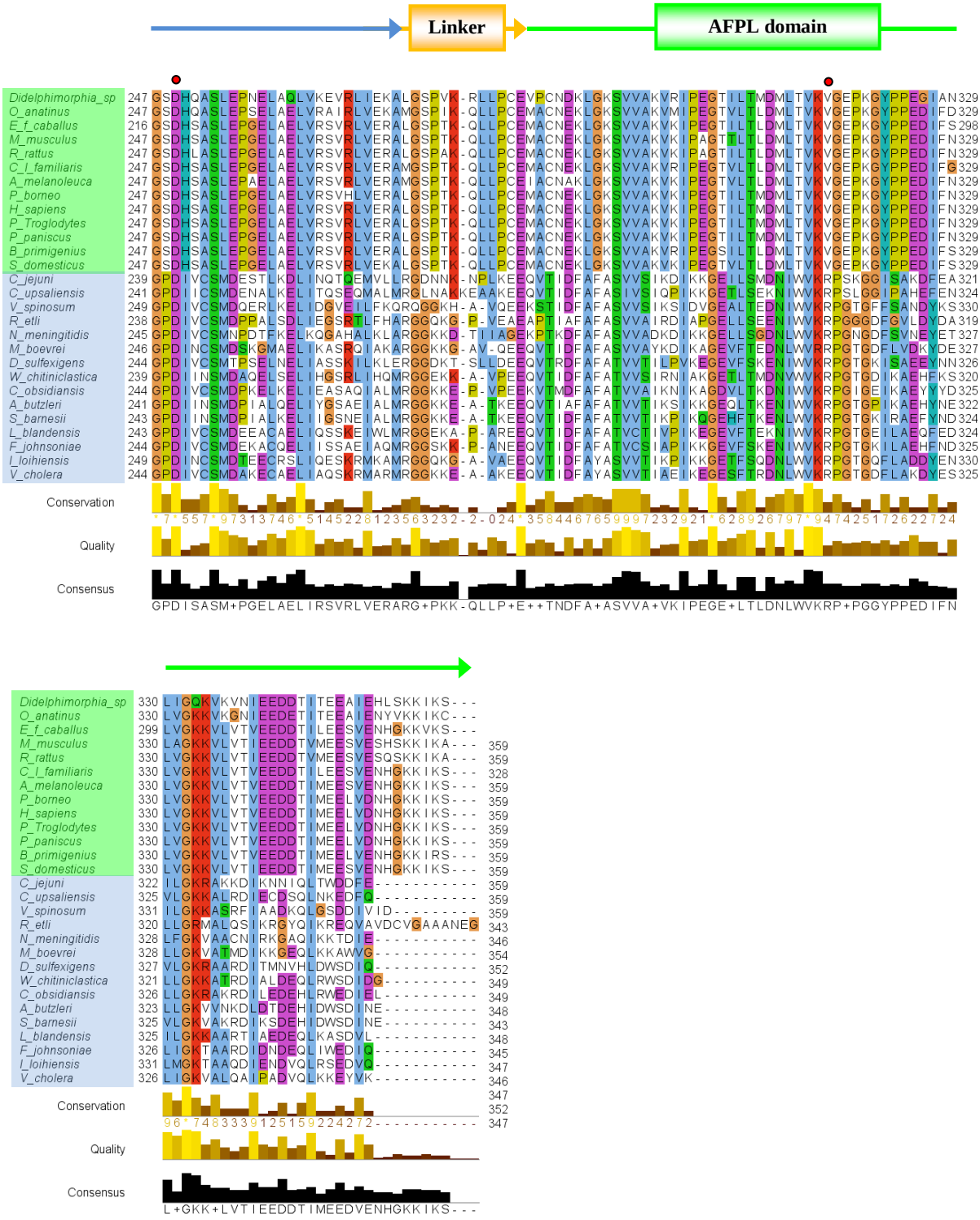


Figure 3.2. Multiple sequence alignment of mammalian NeuNAc 9-PS (green) and bacterial NeuNAcS (blue). Domains and active site residues are annotated according to the *Nme*NeuNAcS crystal structure (PDB: 1XUZ), and sequences coloured using the ClustalX colour scheme (Jalview). Sugar-substrate binding residues are marked by red circles, PEP binding residues by yellow circles and metal binding residues by green circles.

From the multiple sequence alignment it is apparent that the residues comprising the catalytic barrel domain are largely conserved between bacterial and mammalian forms. Of the six

active site residues shown to interact with PEP in the solved *Nme*NeuNAcS structure (PDB: 1XUZ), five are conserved across all bacterial and mammalian sequences analysed (Lys-53, Lys-129, Ser-132, Ser-154 and Ser-213). Furthermore the metal coordinating residues (His-215 and His-236) are entirely conserved. The sugar binding residues Tyr-186, Asp-247 and Gln-55 are conserved between bacterial and mammalian sequences, but conserved bacterial sugar binding residues Arg-314 and Asn-74 are conserved as valine and threonine residues respectively in mammals.

While the catalytic barrel domain residues are mostly conserved between bacterial and mammalian forms of this enzyme, there are two regions which display poor sequence homology. The $\beta_2\alpha_2$ loop of the catalytic barrel (residues 57-79 in *Nme*NeuNAcS) is one such region, and displays poor sequence conservation with the corresponding mammalian loop (Figure 3.2). This loop contains a number of small secondary structural elements (one 4 residue α-helix and one 5 residue 3/10 helix), but shall be referred to purely as the $\beta_2\alpha_2$ loop from here onwards. Residues 69-79 represent the flexible portion of the loop.

Interestingly, the bacterial and mammalian $\beta_2\alpha_2$ loop sequences are highly conserved within their own clades (Figure 3.2). This high degree of intra-clade conservation but poor overall conservation suggests that these loop regions may have evolved to provide an alternate function in bacteria and mammalian forms of the enzyme. Given the proximity of the $\beta_2\alpha_2$ loop to the sugar binding site, and the fact that Asn-74 from this loop anchors the C-6 hydroxyl of rManNAc in *Nme*NeuNAcS (Figure 3.3), it could be predicted that this loop plays an important role in selecting for, and stabilising the respective bacterial and mammalian sugar-substrates, which differ only by the C-6 substituent.[32]

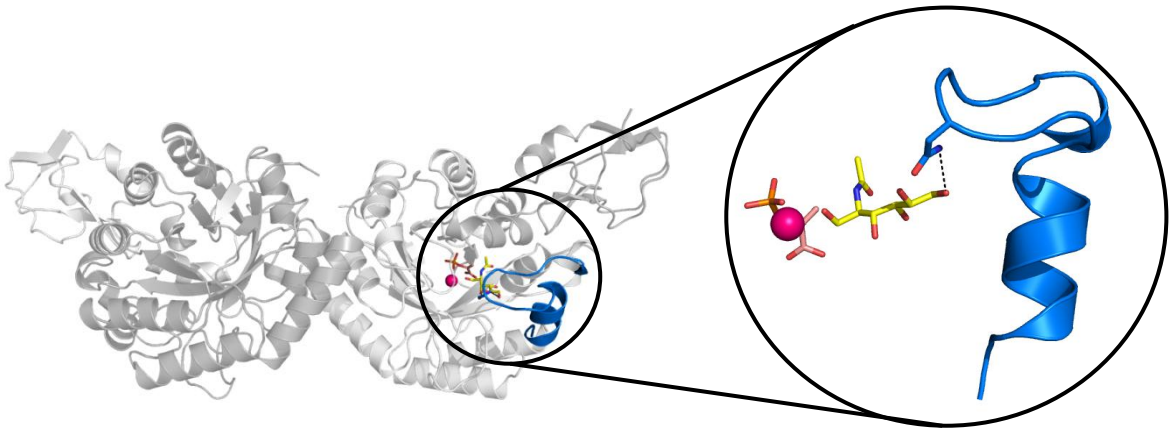


Figure 3.3. The $\beta_2\alpha_2$ loop (blue) of *Nme*NeuNAcS forms direct interactions with the C-6 hydroxyl of ManNAc (yellow) through the side-chain of Asn-74.

What makes this hypothesis more compelling is the abundance of conserved residues with positively charged side-chains within the mammalian loop and flanking regions (Figure 3.4). These residues are largely conserved as non-cationic amino acids in bacterial sequences (Figure 3.2). Given the mammalian sugar-substrate contains a polyanionic C-6 phosphate group which orients toward the $\beta_2\alpha_2$ loop as shown in chapter two, the cationic residues of the loop may provide electrostatic stabilisation of ManNAc 6-P.

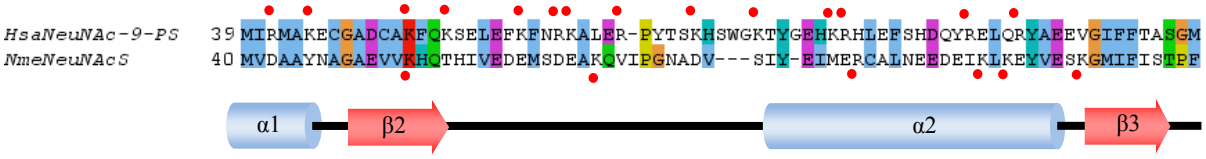


Figure 3.4. Partial sequence alignment of *Hsa*NeuNAc-9-PS and *Nme*NeuNAcS showing the $\beta_2\alpha_2$ loop sequence and major secondary structural elements. Cationic residues are indicated by red circles.

Comparison of the surface electrostatics of the *Hsa*NeuNAc 9-PS model and *Nme*NeuNAcS structure shows a clear difference in the distribution of surface charge across the $\beta_2\alpha_2$ loop region (Figure 3.5). While the $\beta_2\alpha_2$ loop of *Nme*NeuNAcS is predominantly negatively charged at physiological pH, the corresponding region of *Hsa*NeuNAc 9-PS appears to have considerable electropositive surface charge, consistent with the involvement of this loop in binding an anionic phosphorylated substrate.

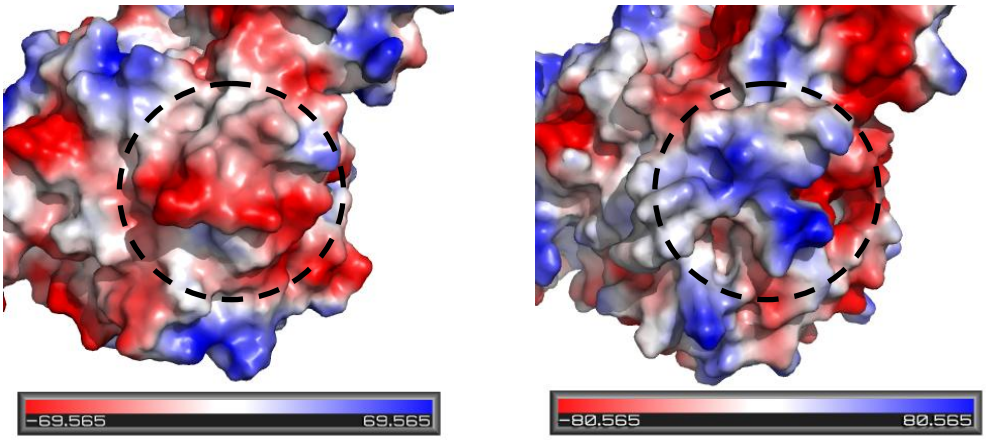


Figure 3.5. PyMOL vacuum electrostatics calculation shows a notable difference in the charge distribution between the $\beta_2\alpha_2$ loop (circled) of *Nme*NeuNAcS (left, PDB: 1XUZ) and modelled *Hsa*NeuNAc 9-PS (right). The $\beta_2\alpha_2$ loop of *Nme*NeuNAcS has considerable negative surface charge (red) whereas the $\beta_2\alpha_2$ loop of *Hsa*NeuNAc 9-PS exhibits a high degree of electropositive charge (blue). Note: this electrostatic calculation is not quantitative and provides an approximate representation of the charge-smoothed surface only.

Within the conserved mammalian $\beta_2\alpha_2$ loop are three residues (Gly-78, Lys-79 & Thr-80), which are reminiscent of a partial "Walker phosphate binding motif" [GXXXXGK(T/S)].[85] From induced-fit substrate docking (Section 2.3) we know that Thr-80 of this motif is a likely phosphate binding residue and thus implicates the $\beta_2\alpha_2$ loop in phosphate binding directly. It seems likely that variability in loop sequence between bacterial and mammalian enzyme dictates the strict sugar-substrate specificity seen in nature.

A second region of poorly conserved sequence within the entire sialic acid synthase family is the flexible linker region (Residues 272-282) which connects the catalytic barrel and C-terminal AFPL domains (Figure 3.2). Furthermore, the AFPL domain itself (Residues 282-349) has relatively low sequence conservation between bacterial and mammalian orthologues. *Hsa*NeuNAc 9-PS and *Nme*NeuNAcS share only 21.8% identity across the combined linker and AFPL region (Residues 272-349) and 22.4% for the AFPL domain alone (Figure 3.6). Comparatively, the catalytic barrel domain (Residues 1-272) of these enzymes share 30.7% identity at the amino acid level. Again, the AFPL domain and linker sequences are almost entirely conserved within mammalian sequences.

It is established that the AFPL domain is essential for the catalytic function and stability of both bacterial and mammalian enzymes.[4, 77] Truncation of the AFPL domain has been shown to totally abolish the activity of both *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS. The conserved bacterial residue Arg-314 from the AFPL domain is known to be essential for catalysis in *Nme*NeuNAcS, and functions to steer the sugar-substrate ManNAc into a catalytically favourable orientation within the active site.[64] The guanidinium side-chain of Arg-314 extends into the active site of one chain, from the AFPL domain of the second chain. We have shown that the conserved mammalian AFPL domain residue Lys-290 plays a similar role, extending into the active site to form specific interactions with the C-6 phosphate group of ManNAc 6-P (Section 2.3). The lack of sequence homology, and observed roles of both bacterial and mammalian AFPL domain residues in sugar-substrate binding, suggests that this C-terminal domain may additionally contribute toward the sugar-substrate selectivity of the bacterial and mammalian sialic acid synthases. Furthermore, the flexible linker region of *Nme*NeuNAcS comes within close proximity of the $\beta_2\alpha_2$ ManNAc binding loop, highlighting the possibility of interplay between these two putative sugar-substrate selectivity elements.[32]

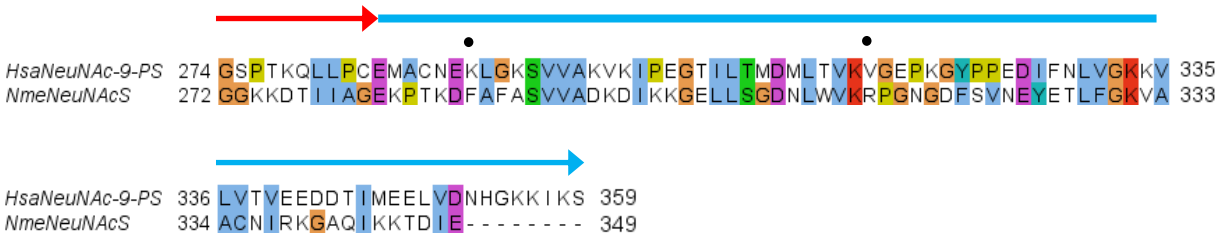


Figure 3.6. Partial alignment of *Hsa*NeuNAc 9-PS and *Nme*NeuNAcS sequences showing the linker region (red) and AFPL domain (blue). Residues implicated in sugar-substrate binding (*Hsa*Lys-290, *Nme*Arg-314) are indicated by black circles.

3.3 Chimera design and generation

Bioinformatic analyses reveals the sequences corresponding to both the $\beta_2\alpha_2$ loop of the catalytic domain, and the AFPL domain/linker are poorly conserved between bacterial NeuNAcS and mammalian NeuNAc 9-PS, but maintain a high degree of conservation within their own clades. This suggests potential involvement in the differentiation and stabilisation of these enzymes alternate sugar-substrates. In an attempt to convert the sugar-substrate specificity of the bacterial enzyme towards that of the mammalian enzyme, a number of chimeric proteins were generated in which the aforementioned structural elements from *Hsa*NeuNAc 9-PS were engineered into the respective positions of the *Nme*NeuNAcS protein scaffold.

3.4 $\beta_2\alpha_2$ loop-swap chimera

In *Nme*NeuNAcS, the $\beta_2\alpha_2$ loop effectively caps the active site, and forms a hydrogen bond with the C-6 hydroxyl of ManNAc via the side-chain of the conserved bacterial residue Asn-74 (Figure 3.3). In mammalian NeuNAc 9-PS, residues of the $\beta_2\alpha_2$ loop have been shown to form similar interactions with the C-6 phosphate group of ManNAc 6-P (Section 2.3). Interestingly, the $\beta_2\alpha_2$ loop of *Hsa*NeuNAc 9-S is two residues longer than *Nme*NeuNAcS, and contains a number of lysine and arginine residues with positively charged side-chains at physiological pH (Figure 3.4). These cationic residues are otherwise absent from the bacterial sequence and may be responsible for providing an electrostatic environment which favours binding of a charged substrate.

Because of the postulated involvement of the $\beta_2\alpha_2$ loop in determining sugar-substrate specificity, a loop-swap chimera was generated in which 14 residues constituting the flexible portion of the *Hsa*NeuNAc 9-PS $\beta_2\alpha_2$ loop (residues 67-80, as predicted by homology

modelling) were engineering into the corresponding position of the *Nme*NeuNAcS $(\beta\alpha)_8$ barrel (Figure 3.7). It was envisaged that insertion of the human $\beta_2\alpha_2$ loop residues into the bacterial enzyme may confer the ability to utilise ManNAc 6-P in catalysis, a novel activity not seen in wild-type bacterial NeuNAcS.[28]

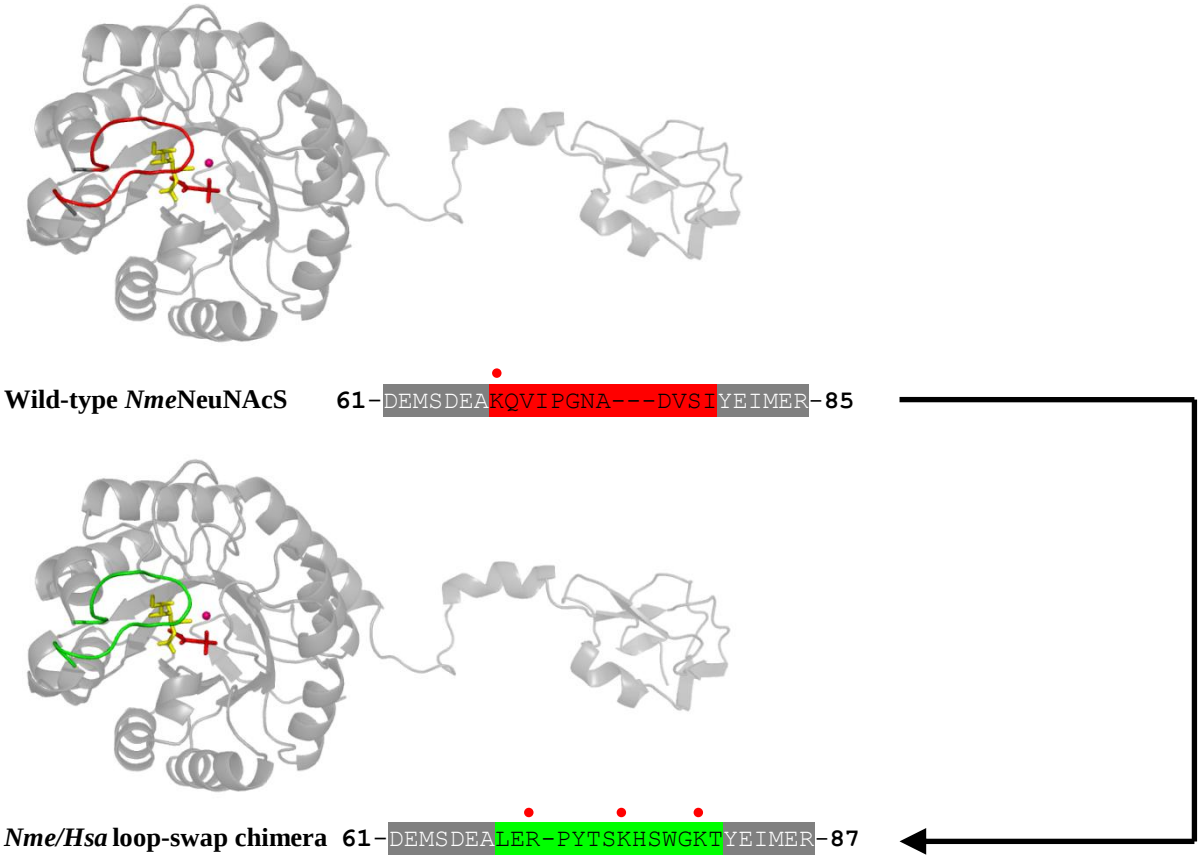


Figure 3.7. A loop-swap chimera was generated in which 12 residues comprising the flexible $\beta_2\alpha_2$ loop of *Nme*NeuNAcS (red) were replaced with 14 analogous residues from the *Hsa*NeuNAc 9-PS $\beta_2\alpha_2$ loop sequence (green). Residues with positively charged side-chains are indicated by red circles.

The gene construct encoding the *Nme/Hsa* loop-swap chimera was successfully generated using a 7-step overlap extension PCR methodology as shown in Figure 3.8. Forward and reverse primers, each encoding half of the 14 *Hsa*NeuNAc 9-PS loop residues to be inserted into the *Nme*NeuNAcS scaffold, were designed and used to amplify a short 'pre-loop' N-terminal fragment and a longer 'post-loop' C-terminal fragment from the *Nme*NeuNAcS wild-type gene (serotype B strain MC58, *neuB*) in the pDEST^{TM}17 vector. The resulting linear fragments were effectively blunt end constructs containing each half of the new loop sequence (Figure 3.8; Steps 1 & 2).

Forward and reverse 'bridging' primers were designed with complementarity to the central 30 base pairs of the human loop sequence, and used to extend each fragment, thereby introducing an overlap of 30 base pairs between the two fragments (Figure 3.8; steps 3 & 4). The overlapping fragments were then extended using forward and reverse gene primers in the final two PCR reactions to generate the complete construct with the 14 human loop residues *in situ* (Figure 3.8; steps 5 & 6).

An additional PCR step using generic Gateway® primers was required to introduce Gateway® 5' and 3' flanking sequences containing an N-terminal TEV protease cleavage site into the final loop-swap gene construct (Figure 3.8; step 7). The forward and reverse gene primers used for all chimera construction steps contain the generic extensions required to facilitate this step. Gel electrophoresis was used at each stage of cloning to ensure successful amplification of the correct constructs.

Following extension with generic Gateway® sequences, the linear construct was ligated into pDONR^{TM}221 entry vector, using the Gateway® protocol outlined in section 6.3.5. The successfully sequence verified pDONR^{TM}221 vector bearing the loop-swap chimera construct was then transformed into *E. coli* One Shot TOP10 cells (Invitrogen) and propagated for plasmid extraction (Section 6.3.8). Purified plasmid from a resulting transformant was ligated into the expression vector pDEST^{TM}17 which encodes an N-terminal His-tag using the protocol described in section 6.3.4. All primers used for chimera cloning are listed in Table 6.3.

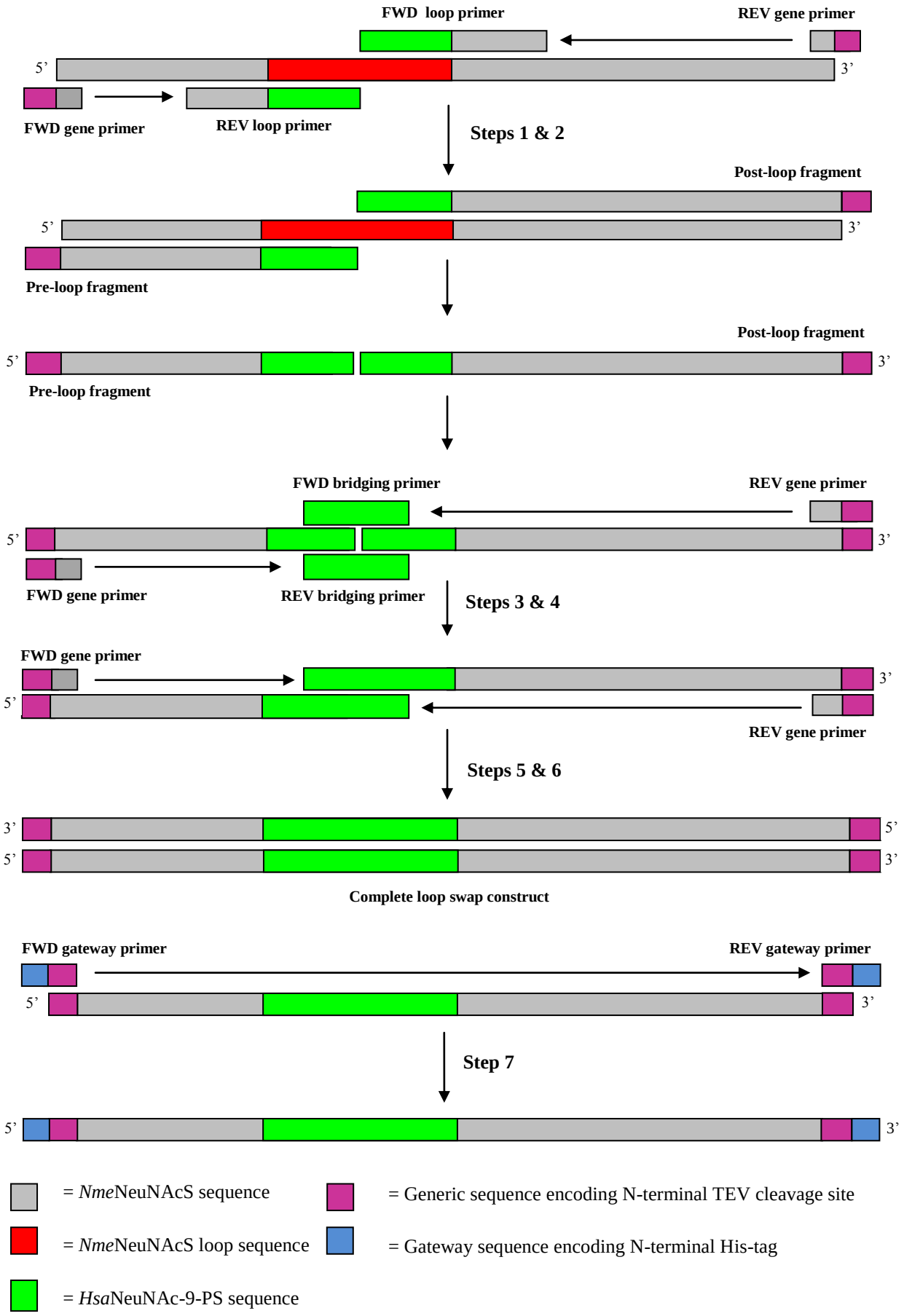
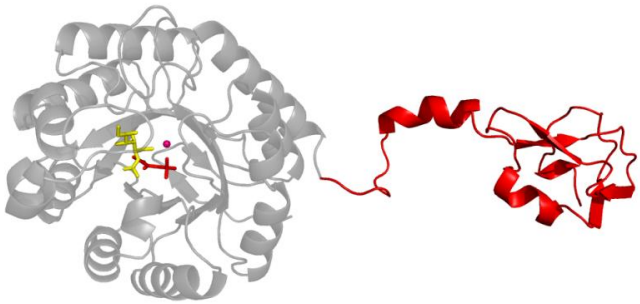


Figure 3.8. Assembly of the linear gene construct encoding the *Nme/Hsa* loop-swap chimera

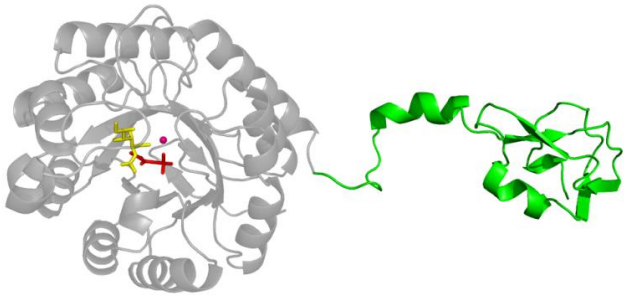
3.5 AFPL domain-swap chimera

Considering the relatively low sequence homology of bacterial and mammalian AFPL domains (22% identity between *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS), and the fact that the AFPL domain contributes residues which interact with sugar-substrates directly in both *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS, it is possible that the AFPL domains are involved in differentiating between and stabilising the respective unphosphorylated or phosphorylated sugars. To this end, an AFPL domain-swap chimera was designed and generated by Dr. Joseph, in which the AFPL domain of *Hsa*NeuNAc 9-PS was engineered into the respective position of the *Nme*NeuNAcS scaffold, in an attempt to confer activity with ManNAc 6-P (Figure 3.9).[86]



Wild-type *Nme*NeuNAcS

261-KQGAHALKLARGGKKDTIIAGEKPTKDFAFASVVADKDIKKGELLLSGDNLWVKRPGNGD
FSVNEYETLFGKVAACNIRKGAQIKKTDIE-349



***Nme/Hsa* AFPL domain-swap chimera**

261-KQGAHALKLARGSPTKQLLPCEMACNEKLGKSVVAKVKIPEGTILTMDMLTVKVGEPKGY
PPEDIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS-359

Figure 3.9. An AFPL domain-swap chimera was generated by Dr. Joseph in which 79 residues comprising the AFPL domain of *Nme*NeuNAcS (red) were replaced with 86 residues from the *Hsa*NeuNAc 9-PS AFPL domain sequence (green). The conserved residue Gly-272 was chosen as the point of interchange, and thus the entire

AFPL domain and linker region was substituted with that from *Hsa*NeuNAc 9-PS. Residues implicated in sugar-substrate binding (*Hsa*Lys-290, *Nme*Arg-314) are indicated by black circles.

The aforementioned AFPL domain-swap chimera was generated using a similar overlap extension PCR method to that described in section 3.4. For full details of the AFPL domain-swap chimera cloning protocol please refer to the thesis of Dr. Dmitri Joseph.[86]

3.6 $\beta_2\alpha_2$ loop-swap/AFPL domain-swap double chimera

Both the $\beta_2\alpha_2$ loop and AFPL domain/linker potentially play roles in sugar-substrate binding and selection. As described previously, the AFPL domain linker region comes within close proximity of the $\beta_2\alpha_2$ loop of *Nme*NeuNAcS.[32] It is therefore possible that these elements are interconnected, and that the structural basis for substrate selection may rely on synergistic operation of both elements. In order to determine if the combination of the $\beta_2\alpha_2$ loop and the AFPL domain/linker functions in substrate selection, a double-chimera incorporating both structural elements was generated. Generation of the construct encoding this double-chimera was achieved using the same overlap extension method described in section 3.4.

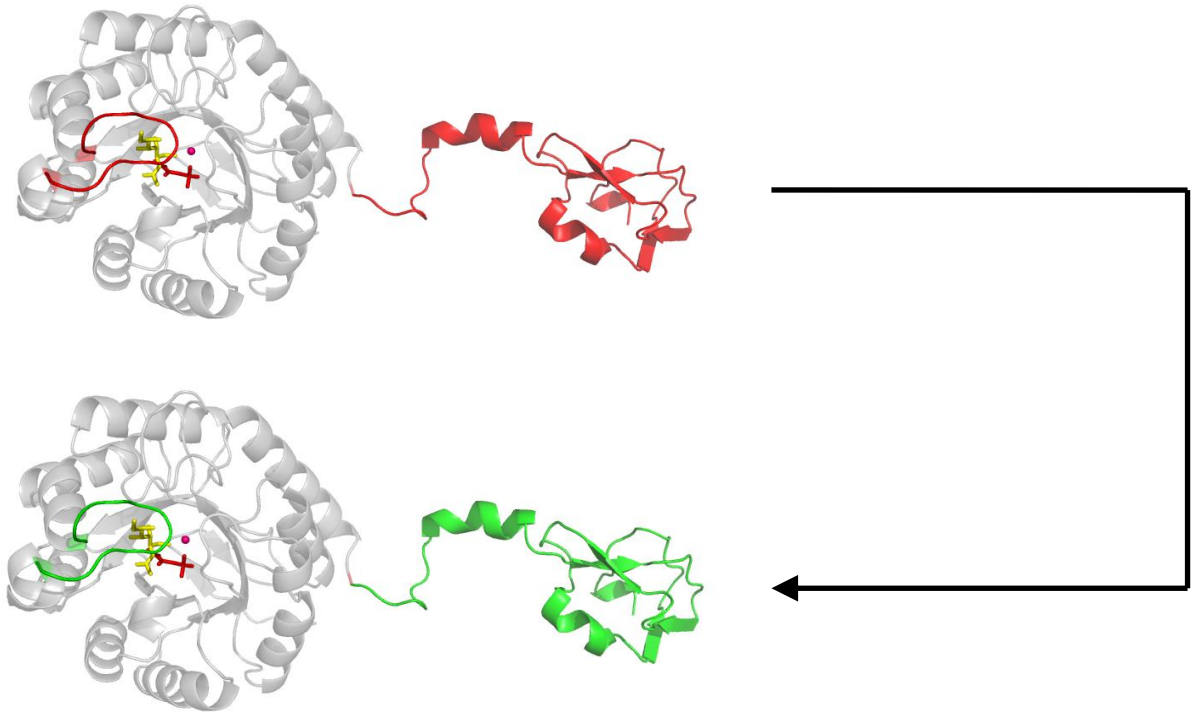


Figure 3.10. A double-chimera was constructed in which both the $\beta_2\alpha_2$ loop and AFPL domain/linker from *Nme*NeuNAcS (red) were replaced with the corresponding residues from *Hsa*NeuNAc 9-PS (green) in an attempt to confer activity with the mammalian substrate ManNAc 6-P. The human loop residues were installed on top of the existing AFPL domain-swap chimera using the original loop-swap chimera primers.

The primers encoding the *Hsa*NeuNAc 9-PS $\beta_2\alpha_2$ loop residues, previously used for generation of the $\beta_2\alpha_2$ loop-swap chimera (Table 6.3), were employed to install the same 14 human enzyme loop residues into the existing AFPL domain-swap chimera construct. This allowed effective insertion of the *Hsa*NeuNAc 9-PS loop sequence on top of the existing AFPL chimera, to produce a final construct with both the human enzyme loop and AFPL domain/linker. All primers used for cloning this chimera are listed in table 6.3.

Generation of the double-chimera construct proceeded as for the $\beta_2\alpha_2$ loop-swap chimera, however the AFPL-domain swap chimera construct (in pDEST^{TM}17) vector was used as the template instead of the wild-type *Nme*NeuNAcS gene (Figure 3.11). Because this template encodes the *Nme*NeuNAcS N-terminal catalytic barrel and *Hsa*NeuNAc 9-PS C-terminal AFPL domain, the combination of forward *Nme*NeuNAcS and reverse *Hsa*NeuNAc 9-PS gene primers was used for extension of linear fragments, and final ligation (Figure 3.11). These gene primers, used for all cloning steps, encode the small generic extensions required for final elongation with Gateway® sequences (Table 6.3).

The complete double-chimera gene construct was successfully amplified and sequence verified. The linear construct was then sub-cloned into the Gateway® expression vector pDEST^{TM}17 using the protocols described in sections 6.3.3 and 6.3.5.

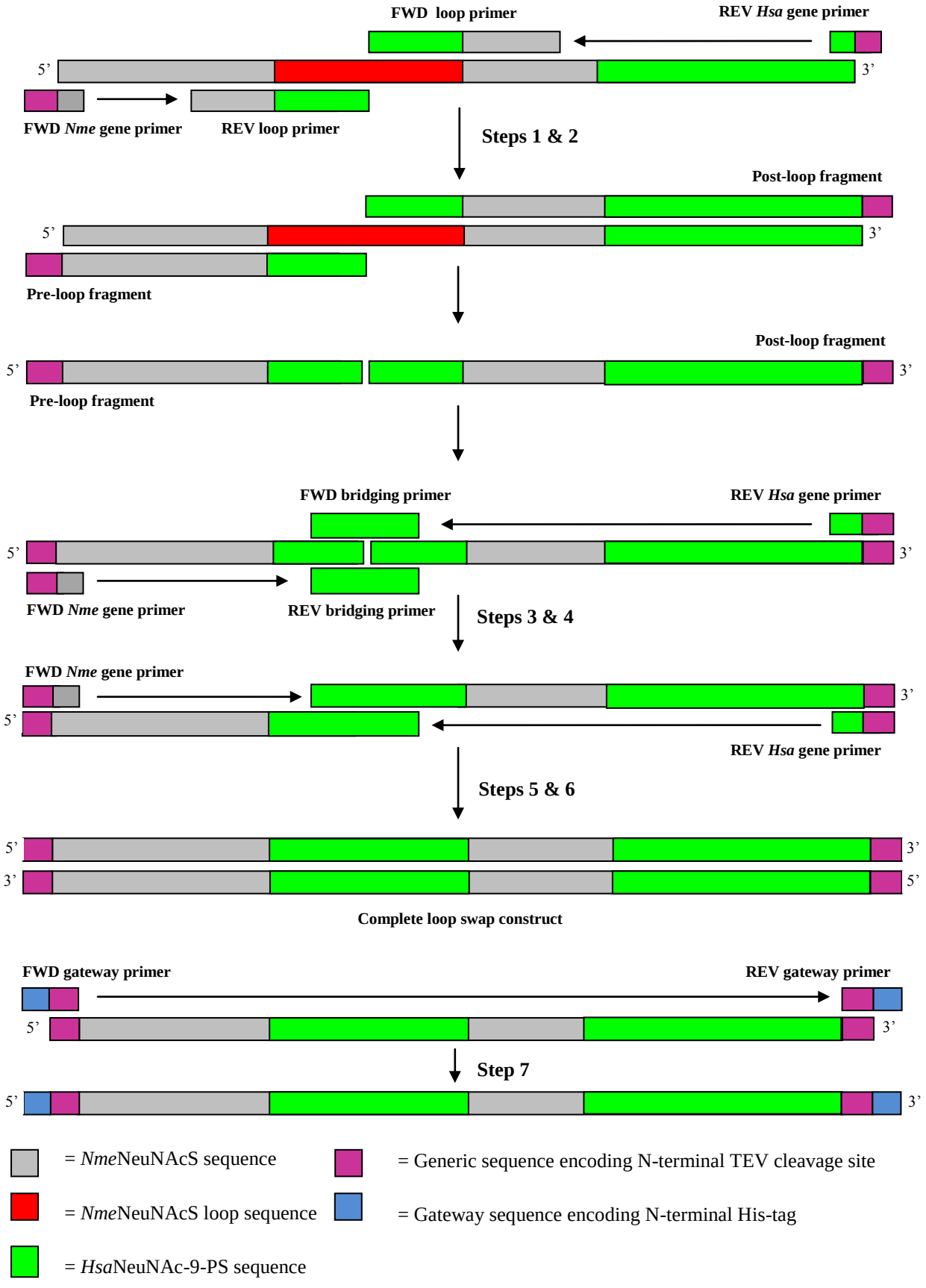


Figure 3.11. Assembly of the linear gene construct encoding the double-chimera using the existing AFPL domain-swap chimera as template.

3.7 Expression and purification of Nme/Hsa chimeras

The pDEST^{TM}17 vectors bearing the $\beta_2\alpha_2$ loop-swap chimera and double-chimera constructs were transformed into *E. coli* BL21 (DE3) Star cells and *E. coli* BL21 (DE3) pBB540/pBB542 (Chaperone 3) cells for expression testing. In order to test inducible expression of the target proteins, an induced culture (containing 0.5 mM IPTG) and non-induced culture of each cell line was grown overnight. Both chimeras were shown (by SDS-PAGE) to be over-expressed following induction with IPTG, and soluble in both cell lines.

The best yield of soluble protein for both chimeras was achieved when expressed in Chaperone 3 cells, and thus this cell line was chosen for upscaled expression and purification. The expressed proteins were purified (Figure 3.12A) using a similar protocol as described for wild-type *Hsa*NeuNAc 9-PS (Section 2.4). Expression and purification of the AFPL domain-swap chimera was completed by Dr. Joseph, as detailed in his thesis.[86]

Following purification, mass spectrometry was used to confirm the correct molecular weights of all chimeras (Appendix B), and circular dichroism (CD) used to ensure the retention of wild-type like secondary structure, and correct folding in solution (Figure 3.12B).

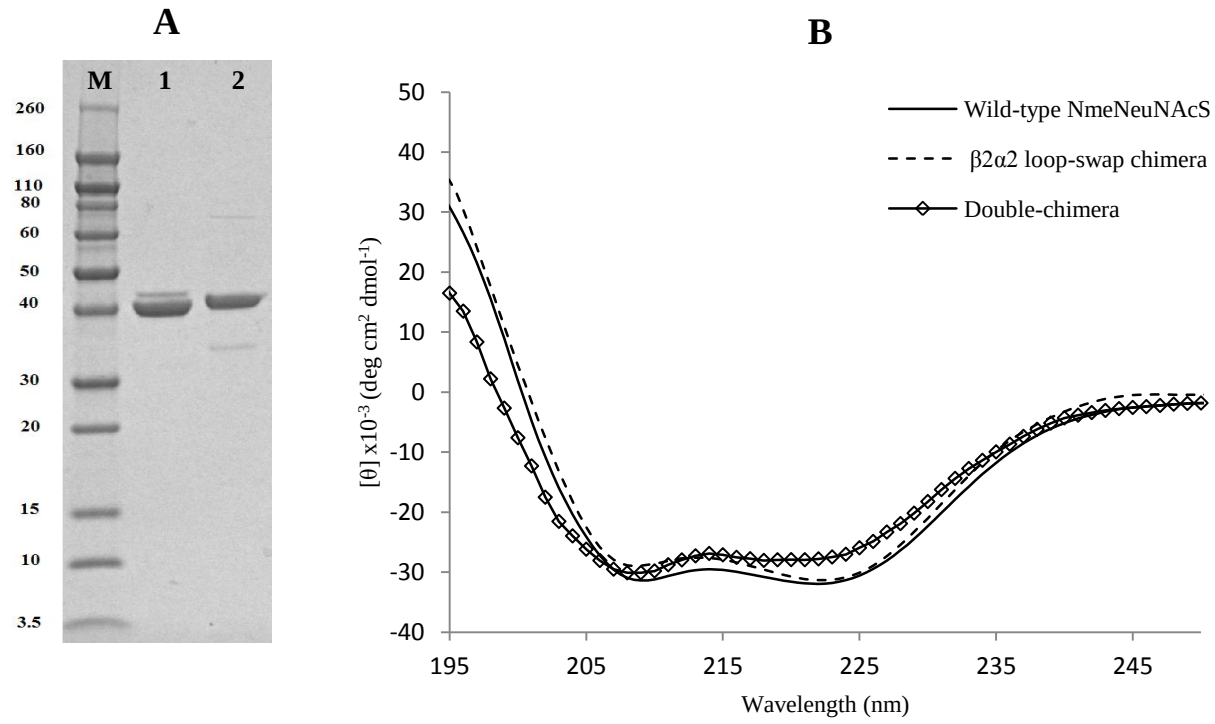


Figure 3.12. A) SDS-PAGE gel showing kDa molecular weight markers (M), purified loop-swap chimera (1) and purified double-chimera (2). **B)** CD spectra of wild-type *Nme*NeuNAcS and chimeras showing retention of native secondary structure and folding.

3.8 Chimera activity assessment with ManNAc and ManNAc 6-P

Following successful purification, the $\beta_2\alpha_2$ loop-swap chimera and double-chimera were assayed for activity with both ManNAc and ManNAc 6-P across a range of conditions, using the continuous UV-Vis assay described in section 6.5.2. The $\beta_2\alpha_2$ loop-swap chimera was shown to retain activity with ManNAc, but at such low levels as to not be quantifiable using this assay. No activity was observable in the presence of the mammalian substrate ManNAc 6-P, even with high concentrations of enzyme (up to 1 mg/mL). No activity was detectable for the double-chimera with either ManNAc or ManNAc 6-P under the experimental conditions used. The AFPL domain-swap chimera was assayed for activity by Dr. Joseph, who concluded there was no observable activity with either ManNAc or ManNAc 6-P, even with protein concentrations as high as 1 mg/mL.[86]

Assaying for low levels of activity by UV-Vis was found to be extremely difficult. Because high enzyme concentrations were required to observe activity, the experimental absorbance values were very high leading to a poor signal to noise ratio, and overall reduction in the accuracy of the assay. Furthermore, it was difficult to differentiate between enzymatic consumption of PEP and non-enzymatic hydrolysis over long periods of time. To circumvent this problem, an NMR assay was developed to screen for very low levels of activity (as described in section 6.5.3). By using NMR, high enzyme concentrations could be used, while retaining the ability to clearly detect product formation. In order to follow enzymatic reactions by 1D proton NMR in an H_2O/D_2O mixture, water suppression using pre-saturation (PRESAT) was required to reduce the otherwise overwhelming solvent signal.

Enzyme activity was determined by monitoring the disappearance of substrate proton resonance, and the proportional appearance of product proton resonances. Specifically, the resonances associated with the α/β forms of the anomeric proton of ManNAc and ManNAc 6-P were monitored throughout the reaction. These proton signals are observable as distinct standalone peaks at 5.0 and 4.9 ppm for α and β protons respectively (Figure 3.13). The axial and equatorial C-3 protons of NeuNAc/NeuNAc 9-P could be observed appearing as the product was formed. In order to verify the assay conditions, wild-type *Nme*NeuNAcS was used as positive control. Total conversion of equi-molar PEP and ManNAc to NeuNAc was observed under the developed assay conditions, validating the methodology (Figure 3.13). Similarly, wild-type *Hsa*NeuNAc 9-PS was used to validate the assay using ManNAc 6-P as

a substrate (not shown). The C-3 proton peaks of NeuNAc 9-P appear in the same position as those of NeuNAc.

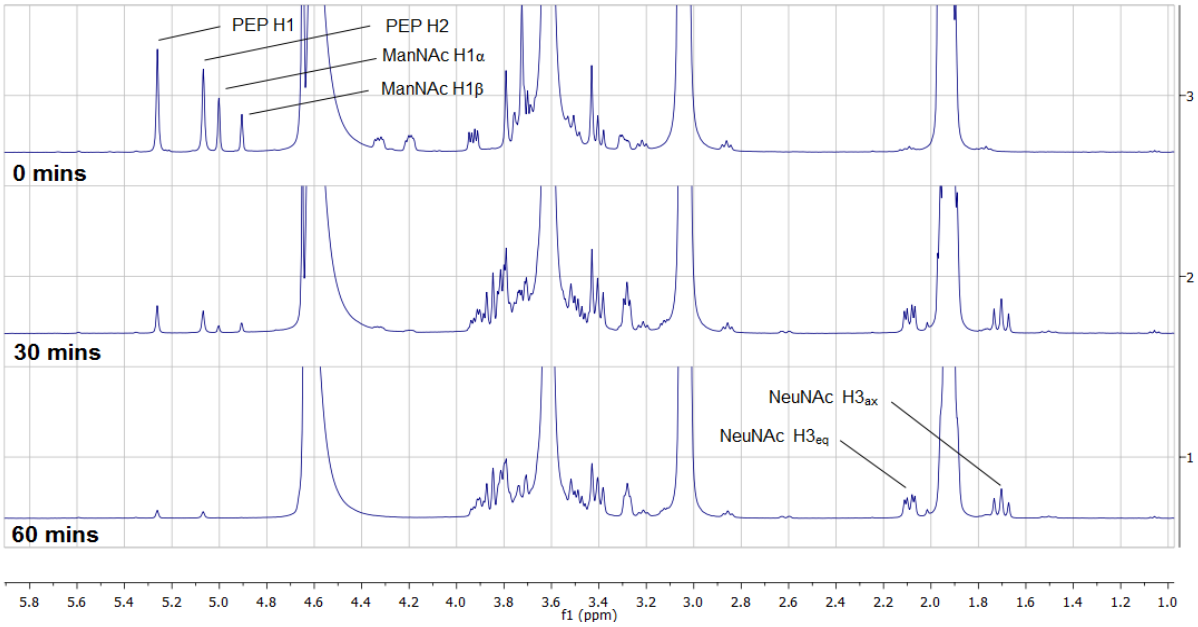


Figure 3.13. PRESAT NMR experiment monitoring the *Nme*NeuNAcS catalysed condensation of 5 mM PEP and 5 mM ManNAc. The appearance of NeuNAc C-3 proton peaks at ~1.7 and 2.1 ppm confirms product formation. Total substrate to product conversion was achieved within ~1hour.

This NMR assay was used to test for low levels of activity for both the $\beta_2\alpha_2$ loop-swap chimera and double-chimera as the lack of activity observable by UV-Vis is not necessarily indicative of complete inactivity, given the constraints of the assay.

The $\beta_2\alpha_2$ loop-swap chimera was confirmed as retaining activity with ManNAc, with distinct peaks consistent with the product NeuNAc, appearing after approximately three hours (Figure 3.14). The rate of turnover however was extremely poor, with approximately 2/3 of substrate converted to product over a period of 24 hours. No product formation was observed when ManNAc 6-P was supplied as the sugar-substrate (Figure 3.15).

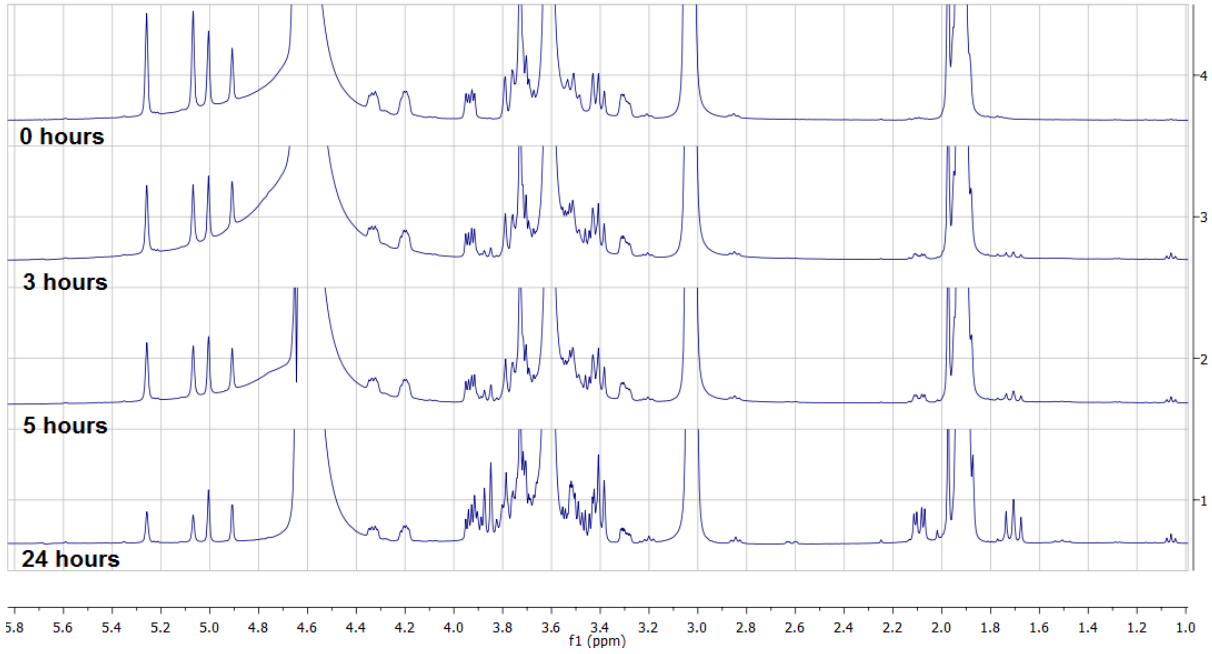


Figure 3.14. PRESAT NMR activity assay for $\beta_2\alpha_2$ loop-swap chimera with 5 mM PEP and 5 mM ManNAc. The appearance of NeuNAc C-3 proton peaks at ~1.7 and 2.1 ppm confirms retention of activity with ManNAc.

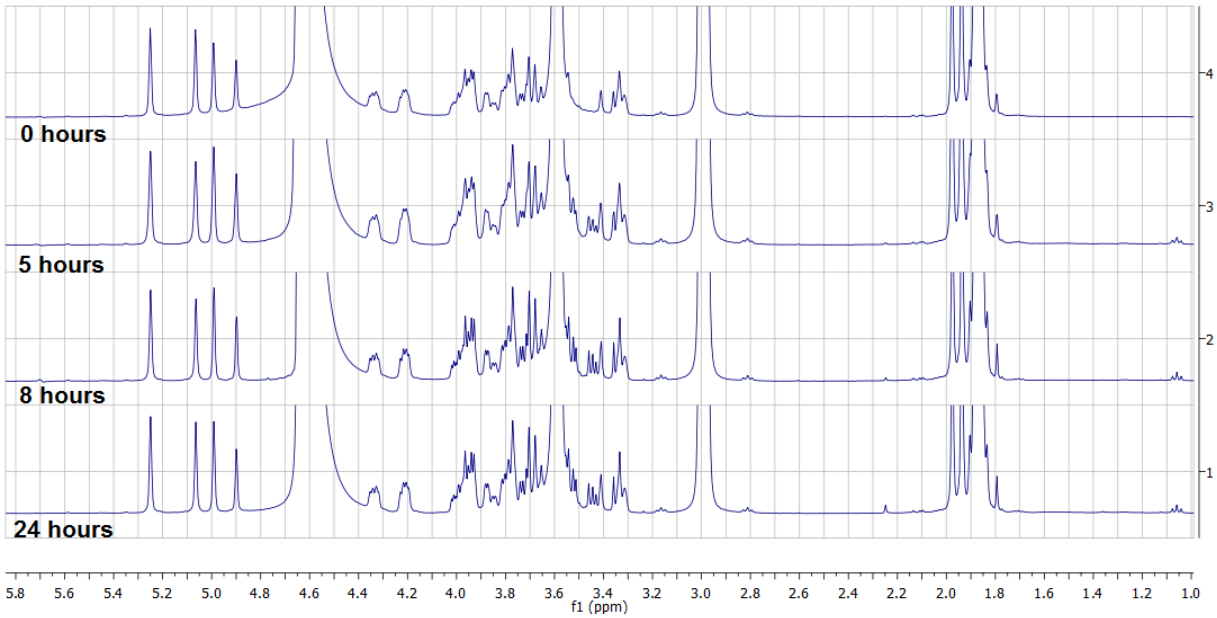


Figure 3.15. PRESAT NMR activity assay for $\beta_2\alpha_2$ loop-swap chimera with 5 mM PEP and 5 mM ManNAc 6-P. No product formation is detectable after 24 hours.

The double-chimera was shown by NMR to be inactive with respect to both ManNAc (Figure 3.16) and ManNAc 6-P (Figure 3.17), with no detectable product formation after 24 hours. It should be noted that in the NMR spectra for the double-chimera with ManNAc 6-P, there is an obvious change in size of the peaks corresponding to the substrates over time (Figure 3.17). This is most likely an artefact of water suppression which may also influence solute

peak sizes from one experiment to the next. The absence of any product resonances provides sufficient evidence to conclude that no substrate turnover is occurring.

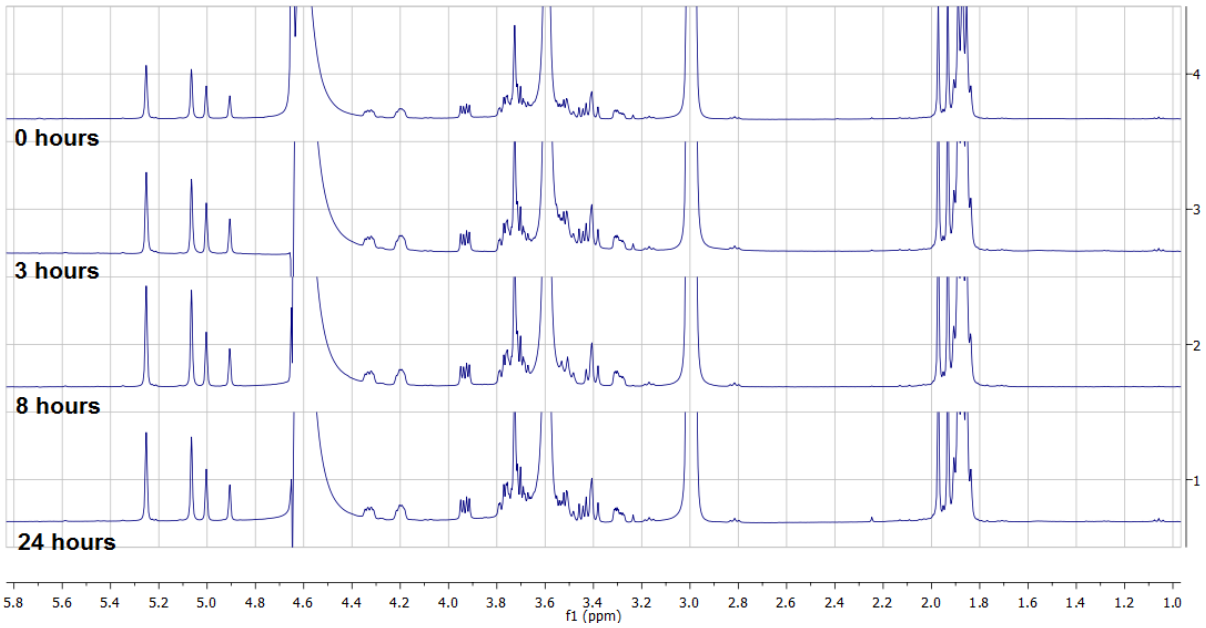


Figure 3.16. PRESAT NMR activity assay for *Nme/Hsa* double-chimera with 5 mM PEP and 5 mM ManNAc. No product formation was detectable after 24 hours.

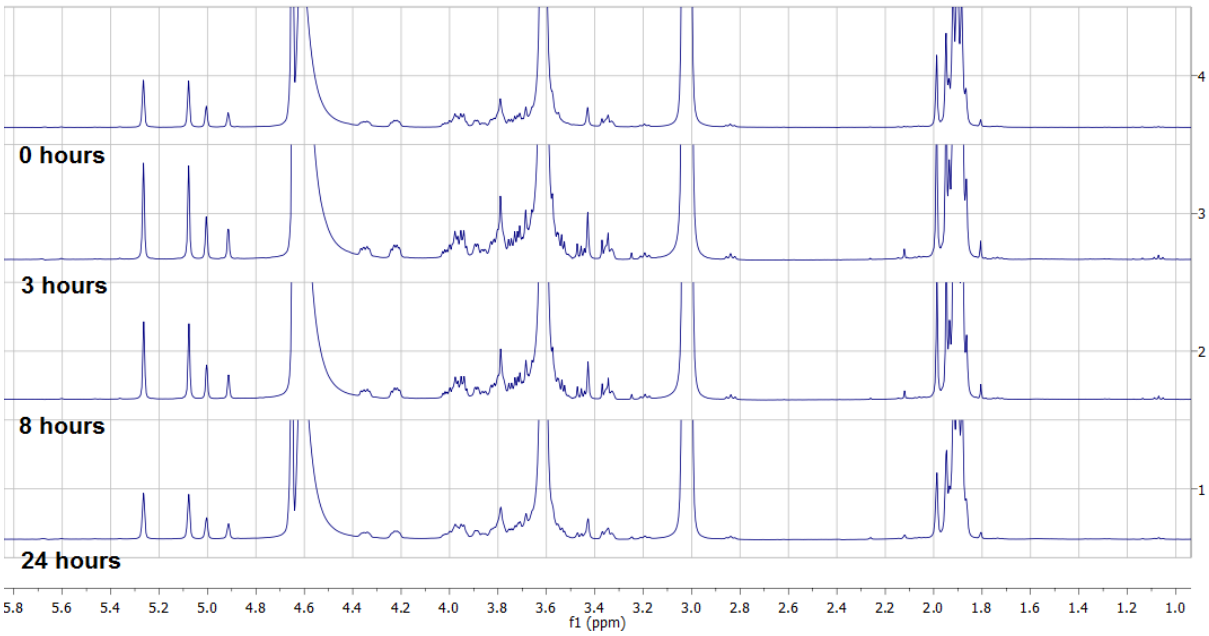


Figure 3.17. PRESAT NMR activity assay for *Nme/Hsa* double-chimera with 5 mM PEP and 5 mM ManNAc 6-P. No product formation was detectable after 24 hours.

3.9 PEP binding studies of Nme/Hsa chimeras

The purified *Nme/Hsa* chimeras were shown to have very poor or entirely abolished activity *in vitro*. Only the $\beta_2\alpha_2$ loop-swap chimera retained any activity, with a slow turnover of ManNAc and PEP observed. Unfortunately, none of the resulting chimeras were able to utilise the mammalian substrate ManNAc 6-P in catalysis. In order to assess the nature of this inactivity, ITC was used to determine whether PEP binding was affected by introduction of the respective structural elements from *Hsa*NeuNAc 9-PS.

The binding interaction of wild-type *Nme*NeuNAcS with PEP was first characterised for comparison with the generated chimeras (Figure 3.18A). Unlike *Hsa*NeuNAc 9-PS, PEP binding is an endothermic binding process for *Nme*NeuNAcS, with an overall ΔH of 5350 ± 140 cal/mol. PEP binding is also associated with a favourable entropic change (TΔS) of -11840 cal/mol. Unexpectedly, when a PEP titration was carried out with the $\beta_2\alpha_2$ loop-swap chimera under identical conditions to that used for wild-type *Nme*NeuNAcS, only a very weak PEP binding event was observable (Figure 3.18B). This binding data could not be fit to a single-site binding model. It is obvious that PEP binding has been significantly impacted by substitution of the $\beta_2\alpha_2$ loop residues, which may account for the considerable reduction in activity observed for this chimera with ManNAc.

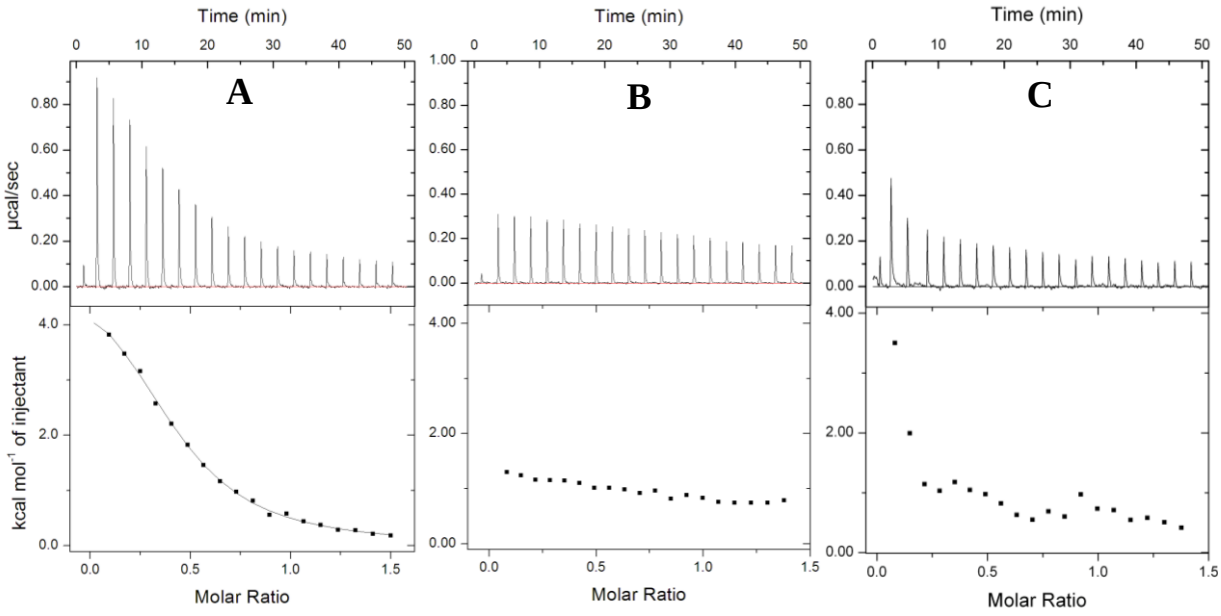


Figure 3.18. Raw ITC thermograms and integrated data for 1 mM PEP titrations into: **A)** 150 µM wild-type *Nme*NeuNAcS, **B)** 150 µM $\beta_2\alpha_2$ loop-swap chimera, and **C)** 150 µM double-chimera.

Similarly, only a very poor PEP binding signal was observed for the double-chimera when a PEP titration was completed under the same conditions as used for wild-type *Nme*NeuNAcS

(Figure 3.18C). PEP binding has apparently been significantly attenuated, potentially via the same mechanism of perturbation as for the $\beta_2\alpha_2$ loop-swap chimera.

While the AFPL domain-swap chimera was initially thought to retain wild-type like PEP binding affinity, it has since been shown that the experimental conditions under which the titration was completed may have resulted in a false-positive binding signal.[86] While this has not yet been corroborated, it is possible that the AFPL domain-swap chimera also has compromised PEP binding affinity.

3.10 Thermal stability of the Nme/Hsa chimeras

DSF was used to determine the relative thermal stabilities of the chimeric proteins with comparison to wild-type *Nme*NeuNAcS in the presence and absence of ligands (Figure 3.19). The unliganded $\beta_2\alpha_2$ loop-swap chimera was shown to have a melt temperature of 46.1 ˚C which is slightly higher than that measured previously for wild-type *Nme*NeuNAcS (42.6 ˚C), while the AFPL domain-swap chimera was shown to have a native melt temperature of 43.6 ˚C.[86]

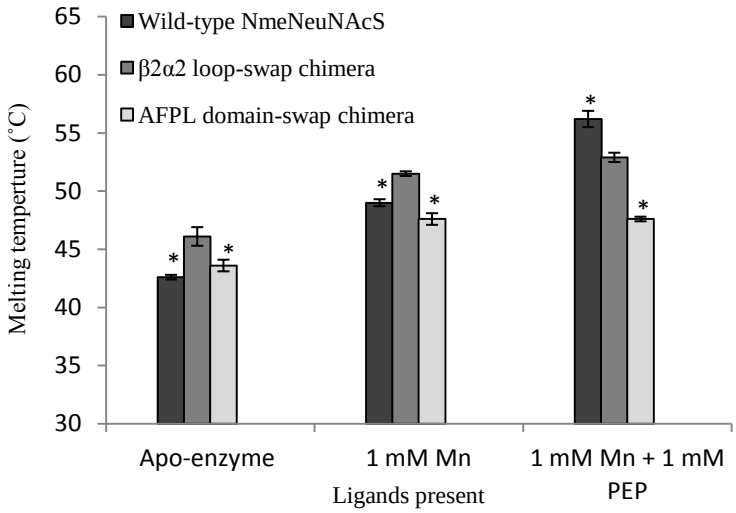


Figure 3.19. Effect of ligands on the thermal stability of wild-type *Nme*NeuNAcS and *Nme/Hsa* chimeras. Data marked with an asterix (*) was collected by Dr. Dmitri Joseph and published in his thesis.[86]

Wild-type *Nme*NeuNAcS is known to undergo significant thermal stabilisation upon Mn^{2+} binding (+6.4 ˚C), and even more-so upon PEP binding (+13.6 ˚C).[77] The $\beta_2\alpha_2$ loop-swap chimera retains stabilisation in the presence of Mn^{2+} (+5.4 ˚C) but unlike wild-type *Nme*NeuNAcS, the presence of PEP affords no additional stabilisation over that already conferred by the metal. The absence of PEP induced stabilisation is consistent with the

observation that this chimera has an attenuated binding affinity for PEP as determined by ITC (Section 3.9). Similarly, the AFPL domain-swap chimera is stabilised in the presence of Mn^{2+} (+4 ˚C), but no additional stabilisation is conferred by PEP. This supports the previously stated hypothesis that this chimera may have a reduced capacity for PEP binding.

Unfortunately, thermal stability data could not be obtained for the double-chimera due to unresolved technical problems with the DSF thermal cycler.

3.11 Structural studies

Analytical gel filtration was used to determine the native molecular masses of the *Nme/Hsa* chimeras (Figure 3.20). The gel filtration column was initially calibrated with a range of known molecular weight standards, and semi-logarithmic plots of elution volume/void volume (V_e/V_o) versus molecular weight were then used to calculate the molecular masses of the chimeras. The $\beta_2\alpha_2$ loop-swap chimera eluted from the column as a single species at 13.7 mL, giving a calculated molecular mass of 85.8 kDa (Table 3.1). This value is closest to that of the theoretical dimeric mass (78.1 kDa), suggesting retention of the wild-type multimeric state under these conditions. The His-tagged double-chimera similarly elutes as a single species at 13.8 mL giving a calculated molecular mass of 73.8 kDa, again closest to that of the theoretical dimer (86.8 Da).

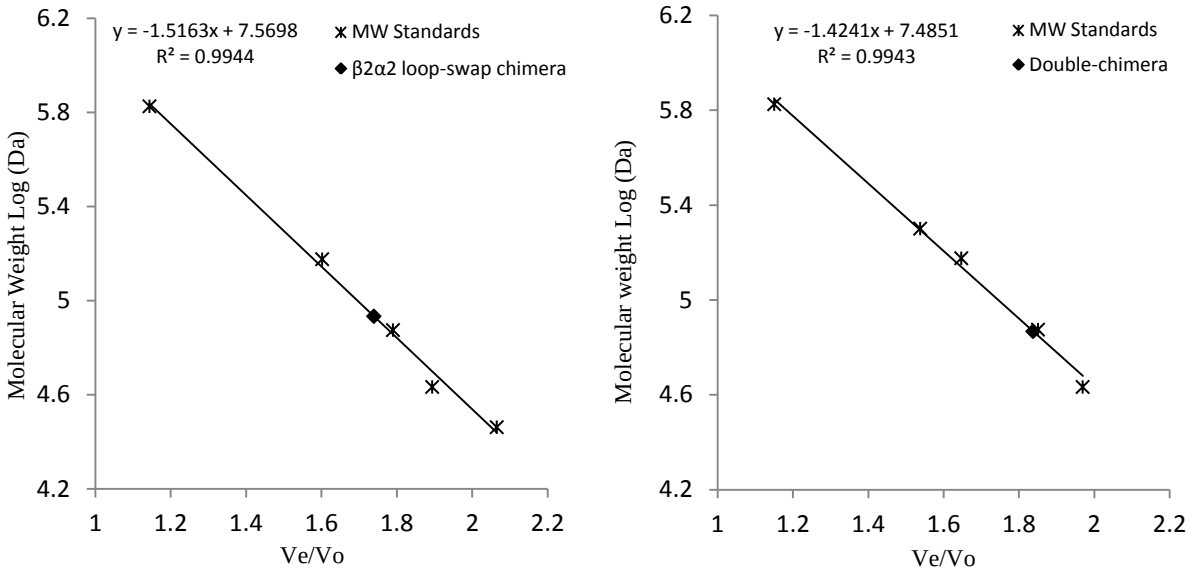


Figure 3.20. Semi-logarithmic plots showing elution volume of target proteins relative to molecular weight standards. Standards used for calibration include at least five of the following: thyroglobulin (669 kDa), beta amylase (200 kDa), alcohol dehydrogenase (150 kDa), ovalbumin (43 kDa), conalbumin (75 kDa) and carbonic anhydrase (29 kDa).

Table 3.1. Experimental parameters determined for *Nme/Hsa* chimeras by analytical gel filtration

	Elution volume (mL)	Elution vol/void vol	Calculated molecular mass (Da)	Theoretical dimeric mass (Da)	Error (%)
$\beta_2\alpha_2$ loop-swap chimera	13.7	1.74	85800	78138	8.9
Double-chimera (His-tagged)	13.8	1.84	73800	86759	14.9

The AFPL domain-swap chimera was also shown to exist as a single dimeric species by analytical gel filtration.[86] The quaternary structure of this chimera was then investigated further by analytical ultracentrifugation (AUC).[86] Sedimentation velocity experiments revealed that a concentration dependant equilibrium exists between monomeric and dimeric species, with dissociation of the dimer observed at concentrations less than 1 mg/mL.[86] Analytical gel filtration was completed at concentrations above this threshold, and therefore only the dimeric species was observed. Wild-type *Nme*NeuNAcS is known to form a stable dimer even at very low concentrations and thus Dr. Joseph concluded that exchange of the AFPL domains had somehow destabilised the native quaternary structure.[86]

Small-angle X-ray scattering (SAXS) was used to verify the quaternary structure of these chimeras in solution. Scattering data was collected for the $\beta_2\alpha_2$ loop-swap chimera in the presence of 1 mM $MnCl_2$ and 1 mM PEP (Figure 3.21A). Scattering data for both the unliganded and PEP bound double-chimera was also obtained (Figure 3.21B).

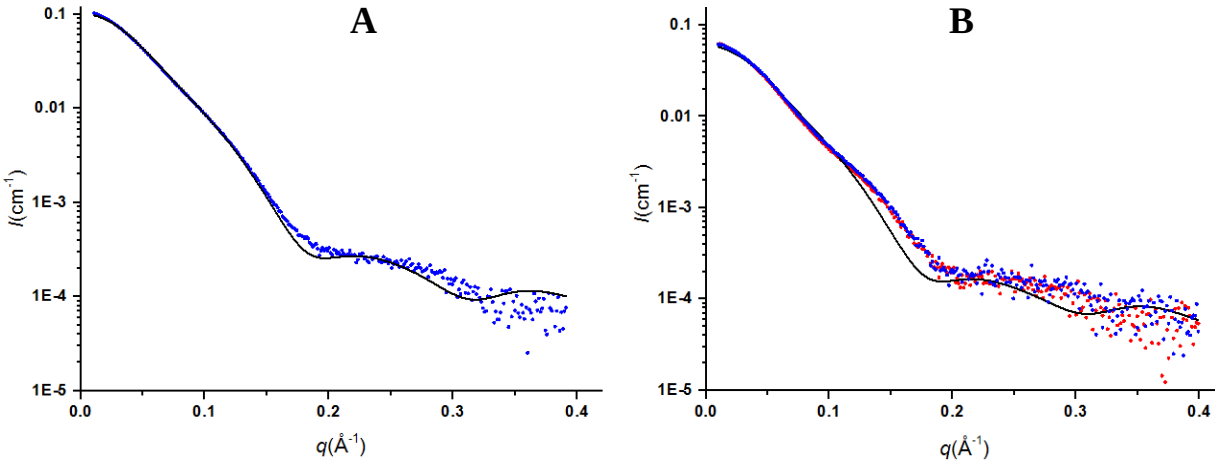


Figure 3.21. A) SAXS profile of $\beta_2\alpha_2$ loop-swap chimera in the presence of 1 mM Mn^{2+} and 1 mM PEP (blue) fit to the theoretical scattering profile of *Nme*NeuNAcS (black line). **B)** SAXS profiles of ligand-free (red) and PEP bound (blue) double-chimera fit to the theoretical scattering profile of *Nme*NeuNAcS (black line).

The SAXS profile of the ligand bound $\beta_2\alpha_2$ loop-swap chimera is in reasonable agreement with the theoretical scattering predicted for ligand bound *Nme*NeuNAcS based on crystallographic coordinates (PDB: 1XUZ), with a chi-square value of 1.725. This indicates probable retention of the domain-swapped arrangement seen for wild-type *Nme*NeuNAcS. The scattering profile of liganded $\beta_2\alpha_2$ loop-swap chimera resembles more closely that of ligand bound wild-type *Nme*NeuNAcS than the apo-enzyme, but lacks the distinct troughs at 0.2 and 0.32 $Å^{-1}$ which are characteristic of the liganded wild-type enzyme.[77] This may suggest the enzyme adopts a less globular conformation, possibly associated with reduced affinity for PEP. As previously described, PEP binding induces a conformational shift toward a more compact species, which facilitates subsequent sugar-substrate binding.[77] The structural parameters derived from SAXS indicate the presence of a dimer under these experimental conditions, consistent with the results of analytical gel filtration (Table 3.2).

The SAXS data obtained for the double-chimera similarly demonstrates dimerisation under the aforementioned experimental conditions. The SAXS derived porod volume for the double-chimera gives a calculated molecular mass of approximately twice the monomeric

mass, and furthermore the radius of gyration (R_g) values are consistent with those calculated for *Nme*NeuNAcS which is known to form a stable dimer in solution (Table 3.2).[86] Both apo and liganded forms show poorer correlation to the predicted *Nme*NeuNAcS profile than observed for the $\beta_2\alpha_2$ loop-swap chimera (Chi-square values of 2.8 and 2.1 respectively). This is unsurprising given the extent of structural disparity from the original wild-type enzyme. While these fits are poor, the improvement of chi-square value, and decrease in the enzymes radius of gyration (R_g) upon PEP addition may suggest a minor shift in the average structural conformation, to something more representative of the liganded wild-type enzyme.

Table 3.2. Experimental parameters obtained by small angle X-ray scattering (SAXS).

Structural parameters	$\beta_2\alpha_2$ loop-swap chimera + $MnCl_2$ + PEP	Double-chimera + $MnCl_2$	Double-chimera + $MnCl_2$ + PEP
$I(0)$ (cm^{-1}) [from P(r)]	0.11 ± 0.01	0.06 ± 0.01	0.06 ± 0.01
R_g (Å) [from P(r)]	34.7 ± 1	35.5 ± 1	34.2 ± 1
$I(0)$ (cm^{-1}) (from Guinier)	0.11 ± 0.01	0.06 ± 0.01	0.06 ± 0.01
D_{max} (Å)	118 ± 6	124 ± 6	118 ± 6
R_g (Å) (from Guinier)	33.8 ± 0.3	35.4 ± 0.4	33.7 ± 0.6
Porod volume estimate ($Å^3$)	112500 ± 11000	118300 ± 12000	109500 ± 11000
Dry dimeric volume calculated from sequence ($Å^3$)	92990	96190	96190
Calculated monomeric molecular mass from sequence (Da)	38797	39751	39751
Molecular mass from Porod volume (Da) (from MoW)	93350	98210	90880

The solution structure of the AFPL domain-swap chimera was assessed by SAXS in the presence of Mn^{2+} and PEP.[86] Dr. Joseph reported a good fit of the experimental scattering data to the theoretical scattering of the *Nme*NeuNAcS crystal structure, with a chi-square value of 1.56.[86] Furthermore, the SAXS profile of the AFPL domain-swap chimera was more similar to experimental scattering obtained for PEP bound *Nme*NeuNAcS, than for unliganded *Nme*NeuNAcS. This may suggest the retention of at least some degree of PEP binding ability, and the ability to undergo (to some extent) the conformational change associated with PEP binding in the wild-type enzyme.[86] The liganded AFPL domain-swap chimera scattering is in better agreement with the scattering profile of liganded wild-type *Nme*NeuNAcS, than the liganded double-chimera. This suggests that the extra substitution of

the $\beta_2\alpha_2$ loop may have had an additional effect on PEP binding over and above the effect of AFPL domain substitution alone.

3.12 Attempted crystallisation

The *Nme/Hsa* chimeras were screened for crystals using the conditions previously described for *Nme*NeuNAcS,[32] and with commercially available screening conditions (Molecular Dimensions PACT, JCSG and Clear Strategy I & II). Protein was screened at a range of concentrations (2, 5 10 mg/mL) in the presence and absence of ligands, however crystals of diffraction quality were not obtained.

3.13 Discussion

Bacterial and mammalian NeuNAc biosynthesis differs primarily in the sugar-substrates which undergo condensation with PEP in the central step of the pathway. Bacterial NeuNAcS catalyses the condensation of ManNAc with PEP, whereas the analogous mammalian enzyme, NeuNAc 9-PS, uses the C-6 phosphorylated derivative ManNAc 6-P. Intriguingly, the aforementioned enzymes are wholly specific for their own substrates, which differ only by the substituent at C-6. As described in chapter two, the human enzyme gains a considerable catalytic advantage over its bacterial counterpart by utilising a phosphorylated substrate with an inherently higher affinity.

Multiple sequence alignment of bacterial and mammalian orthologues identified two distinct structural elements which have reduced sequence homology between bacterial and mammalian forms, but are highly conserved within their own clades. The first of these, the $\beta_2\alpha_2$ loop of the catalytic domain, is known to contribute sugar-substrate binding residues in both *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS (Section 2.3).[32] Interestingly, the human enzyme's loop sequence contains a much higher proportion of conserved residues with positively charged side-chains than that of the bacterial loop. This has potential implications in binding and stabilisation of the polyanionic phosphate group, which is present only in the mammalian substrate. The second identified region is the AFPL domain and the short linker which connects it to the catalytic barrel. The sequence of this unique C-terminal domain is poorly conserved between bacterial and mammalian forms, and similarly contributes residues known to participate directly in sugar-substrate binding.[64] The linker region, consisting of approximately 10 residues, comes within close proximity of the aforementioned putative sugar-substrate binding loop upon formation of the functional domain swapped homo-

dimer.[32] This proximity alludes to potential cross-talk between these two putative sugar-substrate selectivity elements.

A number of chimeric proteins were generated in order to probe the importance of these clade-specific sequences, and as an attempt to convert the activity of a bacterial NeuNAcS to that of mammalian NeuNAc 9-PS by conferring the ability to utilise the phosphorylated substrate ManNAc 6-P in catalysis. The first of these involved the introduction of 14 residues from the $\beta_2\alpha_2$ loop of *Hsa*NeuNAc 9-PS into the corresponding position of the *Nme*NeuNAcS catalytic domain. Dr. Joseph then generated a complementary chimera in which the AFPL domain of *Hsa*NeuNAc 9-PS was fused to the catalytic domain of *Nme*NeuNAcS.[86] Finally, a double-chimera was generated in which both $\beta_2\alpha_2$ loop and AFPL domain sequences from the human enzyme were installed into the *Nme*NeuNAcS scaffold simultaneously.

The aforementioned chimeras were expressed as soluble proteins and purified to homogeneity. Furthermore, all were shown to be correctly folded in solution and retain wild-type like dimeric quaternary structure. Unfortunately none of the resulting chimeras had adopted the ability to utilise ManNAc 6-P in catalysis. The loop-swap chimera was unique in the fact that it retained low levels of activity with the native substrate ManNAc, while the remaining two chimeras were entirely inactive with ManNAc.

Gross reduction of activity with ManNAc upon substitution of the $\beta_2\alpha_2$ loop is not at all unexpected, as this substitution removes the functional side-chain of Asn-74, a residue known to bind the C-6 hydroxyl of ManNAc in the wild-type enzyme (Figure 3.3).[32] The fact that the loop-swap chimera retains any activity with the bacterial substrate is remarkable, and suggests the substituted human loop sequence is capable of maintaining the ManNAc binding function in part. The conserved human residue Thr-72, predicted to interact with the C-6 phosphate of ManNAc 6-P (Section 2.3), may be able to fulfil the role of Asn-74 in the loop-swap chimera. Because the side-chain of Thr-72 is shorter than that of Asn-74, this interaction may be weaker and/or transient and therefore provide less efficient stabilisation of the sugar-substrate.

Inactivity of the double chimera may arise for the same reasons given the loop is also substituted. Furthermore, substitution of the *Nme*NeuNAcS AFPL domain with that of *Hsa*NeuNAc 9-PS is likely to abolish activity with ManNAc through loss of the critical residue Arg-314.[64] The corresponding human residue is a conserved valine, which could provide no such function as that described for Arg-314. Of course it is likely that other, more

global, factors resulting from substitution of these structural elements are also implicated in loss of activity. A subtle change in protein dynamics for example, could potentially destroy the catalytic function of these modified enzymes.

The inability of these chimeras to successfully turnover the mammalian substrate ManNAc 6-P may be due to a number of factors. Substitution with the putative human sugar-binding loop alone is apparently not sufficient to confer activity with ManNAc 6-P. Only 14 residues (of the flexible portion of the loop) were substituted and therefore additional mammalian structural elements which may facilitate functional positioning of the loop, are absent in the resulting chimera. Substitution of a larger portion, perhaps incorporating the entire $\beta_2\alpha_2$ loop, may increase the likelihood of ManNAc 6-P binding. Incorporation of the human AFPL domain in addition to the $\beta_2\alpha_2$ loop sequence should be expected to provide a greater chance of conferring activity with ManNAc 6-P, given what we know about the importance of Lys-290 from the AFPL domain (Section 2.3). This combination of human protein elements is however apparently still insufficient to prompt catalysis with the human substrate.

The leading cause of loss or reduction of activity is almost certainly the inadvertent perturbation of PEP binding which was observed by ITC and/or DSF for all three chimeras. This confounding result makes it difficult to delineate the effect of substituted sequence elements on activity. Perhaps the chimeras would exhibit novel activity with ManNAc 6-P if only PEP binding was retained? For the loop-swap chimera this is probably not the case given retention of activity with ManNAc, however for the AFPL domain-swap and double-chimera proteins this possibility should not be excluded.

There is no obvious connection between the $\beta_2\alpha_2$ loop and the conserved PEP binding residues, however previously reported molecular-dynamics simulations demonstrated that PEP binding results in a reduction of the flexibility of the $\beta_2\alpha_2$ loop of *Nme*NeuNAcS.[77] As this loop occupies a position which partially occludes entrance to the active site, change in the sequence here may alter the sampled conformationals of the loop, thus preventing access of substrates. The relatively poor sequence identity between *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS may mean the human loop sequence is largely incompatible with the bacterial scaffold, and unable to undergo the same range of motion as that of the wild-type enzyme. Repeating the molecular-dynamics simulations conducted for *Nme*NeuNAcS, with the human enzyme loop residues *in situ*, may expose any change in flexibility of the loop that could be responsible for occlusion of PEP from the active site.

Thermal stability data suggests that substitution of the bacterial AFPL domain also affects PEP binding. In *Nme*NeuNAcS, a hydrogen bonding network links PEP binding to rearrangement of the AFPL domain, a process thought to be essential for establishing a competent catalytic site.[77] Only one of the AFPL domain residues involved in the bacterial hydrogen bonding network (Glu-282) is retained in the AFPL domain-swap chimera as this glutamate is also conserved in mammals. Within this hydrogen bonding network is Asn-184, which is known to bind the phosphate group of PEP.[32] Disruption of the hydrogen bonding network through introduction of the human AFPL domain may prevent Asn-184 from maintaining a functional PEP binding orientation within the active site.

Loss of interactions between the AFPL domain and main barrel may also be responsible for destabilisation of the quaternary structure of the AFPL-domain swap chimera (and by analogy, the double-chimera). The AFPL domain-swap chimera was shown to exist in both dimeric and monomeric states at low concentrations, a property not observed for wild-type *Nme*NeuNAcS, which exists predominantly as a stable dimer.[86] Two key interactions identified between the AFPL domain and catalytic barrel of *Nme*NeuNAcS are lost upon substitution with the human enzyme AFPL domain. Arg-314 and Glu-134 form a salt bridge across the dimer interface, however in the AFPL domain-swap chimera, Arg-314 is substituted by the *Hsa*NeuNAc 9-PS residue Val-316, which is incapable of forming such interaction. Similarly, the *Nme*NeuNAcS residues Asp-296 and Arg-114 are shown to form a high-occupancy interaction across the dimer interface, anchoring the globular portion of the AFPL domain of one chain, to the main barrel of the other.[86] Asp-296 is conserved as Lys-298 in mammals, and thus substitution with the mammalian AFPL domain introduces electropositive repulsion at this point. Loss of the aforementioned dimer interface contacts would result in increased mobility of the AFPL domain and promote dimer dissociation. Site-directed mutagenesis could be used to reinstall these original contacts, thereby stabilising the dimer interface and potentially assisting with the transfer of novel activity.

It is apparent that interconverting the activity of bacterial NeuNAcS is no trivial task. The substitution of large regions of sequence gives rise to a plethora of confounding effects, including unintentional disruption of PEP binding and quaternary structure. A more strategic approach to inter-converting activity might be to remodel the bacterial enzyme's active site residues individually, interchanging conserved bacterial residues to the corresponding conserved mammalian residues. The key sugar-substrate binding interactions identified in chapter two would assist in guiding this process. Alternatively, a directed evolution approach

using random mutagenesis and/or gene recombination would provide a powerful route for conferring novel activity. This would however rely on the development of a selection assay that could be used to screen for ManNAc 6-P positive mutants. The resulting ManNAc 6-P active mutants could then be sequenced to verify the specific nature of the protein modification, however because ManNAc 6-P and ManNAc are part of the same biosynthetic pathway to NeuNAc, it would be difficult to distinguish any novel ability to utilise ManNAc 6-P in catalysis from residual activity with ManNAc.

Mammalian and bacterial orthologues of this enzyme are the products of divergent evolution, having undergone remodelling to accommodate the alternate sugar-substrates that they utilise in catalysis. This remodelling is particularly apparent in the sequences of the $\beta_2\alpha_2$ loop and AFPL domain/linker where amino acid conservation drops significantly between bacterial and mammalian clades. These elements are likely to provide the necessary differentiation between ManNAc 6-P and ManNAc. Because the bacterial biosynthetic pathway to NeuNAc lacks the additional kinase and phosphatase catalysed steps of the mammalian pathway, it could be assumed that bacterial NeuNAcS would not be exposed to ManNAc 6-P and therefore would need not be selective against this metabolite. In reality, many microorganisms express ManNAc kinase, however this functions primarily as a catabolic enzyme.[7, 87] In addition to a key cell surface decoration, NeuNAc is utilised as a rich source of carbon, nitrogen and energy by a number of pathogenic microbes.[7, 88] Sialidase activity facilitates the scavenging of free NeuNAc from mammalian host cells, and subsequent cleavage by specific lyase activity releases ManNAc (Figure 3.22) and pyruvate (not shown).

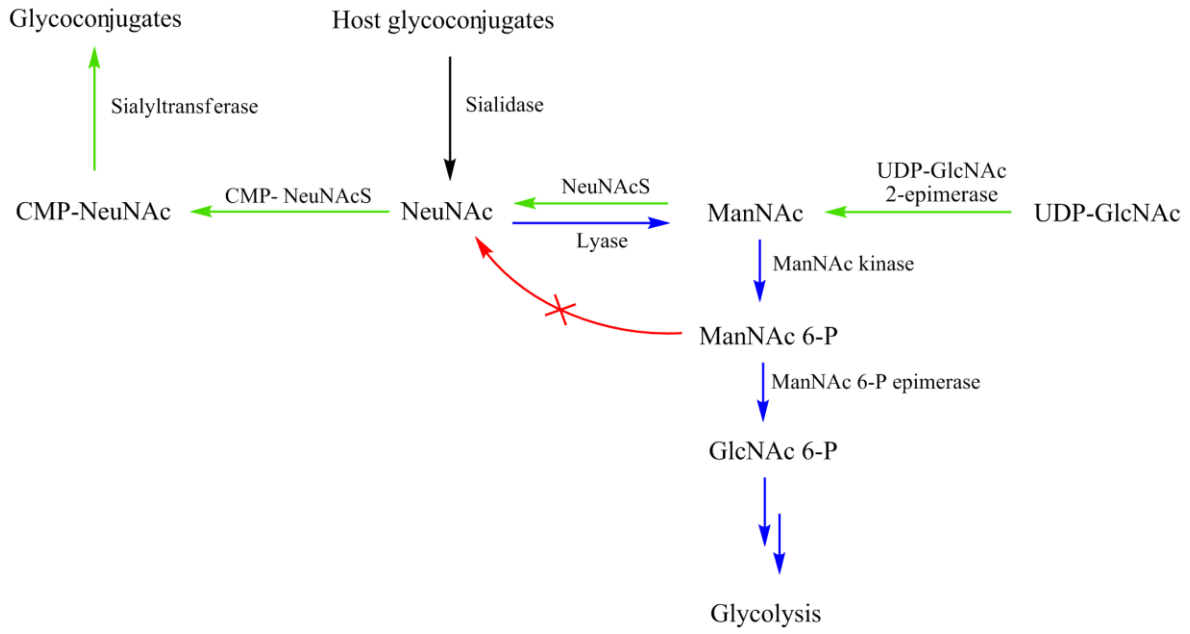


Figure 3.22. The bacterial pathways of NeuNAc biosynthesis (green arrows) and catabolism (blue arrows).

ManNAc is then phosphorylated by ManNAc kinase to give ManNAc 6-P.[7] Unlike in mammalian systems, ManNAc 6-P is destined for catabolism in bacteria, and is epimerised to GlcNAc 6-P which eventually enters the glycolytic pathway.[87] It is not known how bacteria maintain the balance between anabolic and catabolic commitment of ManNAc.[7] What is apparent however is that bacterial NeuNAcS is required to be selective only for non-phosphorylated ManNAc otherwise ManNAc 6-P destined for catabolic pathways would be perpetually returned to the NeuNAc pool (Figure 3.22). By utilising only ManNAc in NeuNAc biosynthesis, ManNAc 6-P becomes a committed catabolic metabolite.

Because the initial UDP-GlcNAc 2-epimerase and kinase steps of mammalian NeuNAc biosynthesis are catalysed by a bi-functional enzyme, metabolite channelling would presumably prevent NeuNAc 9-PS from frequent exposure to unphosphorylated ManNAc.[88] It is therefore possible that this enzyme has lost the ability to utilise ManNAc through active site remodelling to accommodate the phosphorylated substrate, rather than active exclusion of the unphosphorylated amino-sugar. Regardless, by utilising ManNAc 6-P in catalysis, the human enzyme gains improved catalytic efficiency, at the expense of an additional two biosynthetic steps. Because NeuNAc biosynthesis is a core metabolic process in mammals, it is unsurprising that the pathway to NeuNAc has become significantly refined. Bacteria however, utilise NeuNAc expression primarily for host immune evasion, a process not immediately central to survival and proliferation. NeuNAc biosynthesis may therefore be

subject to weaker selection pressures in bacterial systems, and so a less efficient route to NeuNAc is sufficient for meeting cellular demands.

To conclude, a number of sequence elements likely to be involved in sugar-substrate selection by bacterial NeuNAcS and mammalian NeuNAc 9-PS were identified. An attempt to interconvert the sugar-substrate specificity of *Nme*NeuNAcS via the installation of human sequence elements was unsuccessful due to the unforeseen and unintentional disruption of PEP binding. Furthermore, the complex interplay between the AFPL domain, $\beta_2\alpha_2$ loop and catalytic site make this a highly challenging task. The resulting chimeras, while inactive with ManNAc 6-P, may provide promising scaffolds from which to further pursue the transfer of novel activity.

Chapter 4: Sugar-substrate specificity of non-NeuNAcS bacterial sialic acid synthases

4.1 Overview

Microorganisms synthesise a number of unique sialic acid like compounds not observed in higher organisms.[7] Pseudaminic acid (5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-nonulosonic acid) and legionaminic acid (5,7-diacetamido-3,5,7,9 tetradeoxy-D-*glycero*-D-*galacto*-nonulosonic acid) are two such compounds and are known to be important components of bacterial cell surface structures.[11, 46] Heavy post-translational glycosylation of flagellar proteins with the aforementioned 'bacterial sialic acids' is apparently essential for the motility and virulence of a number of pathogenic bacteria, including species of the *Campylobacter* genus.[11, 44]

Pseudaminic acid is biosynthesised in bacteria by condensation of PEP with 2,4-diacetamido-2,4,6-trideoxyaltrose, catalysed by pseudaminic acid synthase (PseS) (Figure 1.8). Similarly, legionaminic acid is the product of condensation of PEP with 2,4-diacetamido-2,4,6-trideoxy-D-*manno*pyranose, a reaction catalysed by yet another member of the sialic acid synthase family; legionaminic acid synthase (LegS) (Figure 1.6). The sugar-substrates of the PseS and LegS catalysed reactions are stereoisomers, and differ by inversion of stereochemistry at C-2, C-4 and C-5. These reactions are analogous to the PEP condensation reactions catalysed by NeuNAcS and NeuNAc 9-PS however the bacterial sialic acid precursor sugars differ from the canonical sialic acid precursor ManNAc through additional *N*-acetylation at C-4 and loss of the C-6 hydroxyl group. Unlike NeuNAcS and NeuNAc 9-PS, which are selective for sugar-substrates based on phosphorylation (or lack thereof) at C-6, these bacterial sialic acid synthases specifically recognise and bind their appropriate sugar-substrates based on variations in stereochemistry and substitution.

The genome of *C. jejuni* contains three genes encoding sialic acid synthase functionality.[11] *Cj1141* has been characterised as a NeuNAcS, *Cj1317* as a PseS and *Cj1327* as a LegS.[42, 48, 54] The presence of genes encoding all three primary sialic acid synthases makes *C. jejuni* an excellent model from which to study the structural and functional variation which defines these enzymes. Elucidation of the molecular basis of the alternate sugar-substrate specificity of these bacterial sialic acid synthases may form the basis for the exploitation of these enzymes for therapeutic advantage. Furthermore, understanding the specific active site

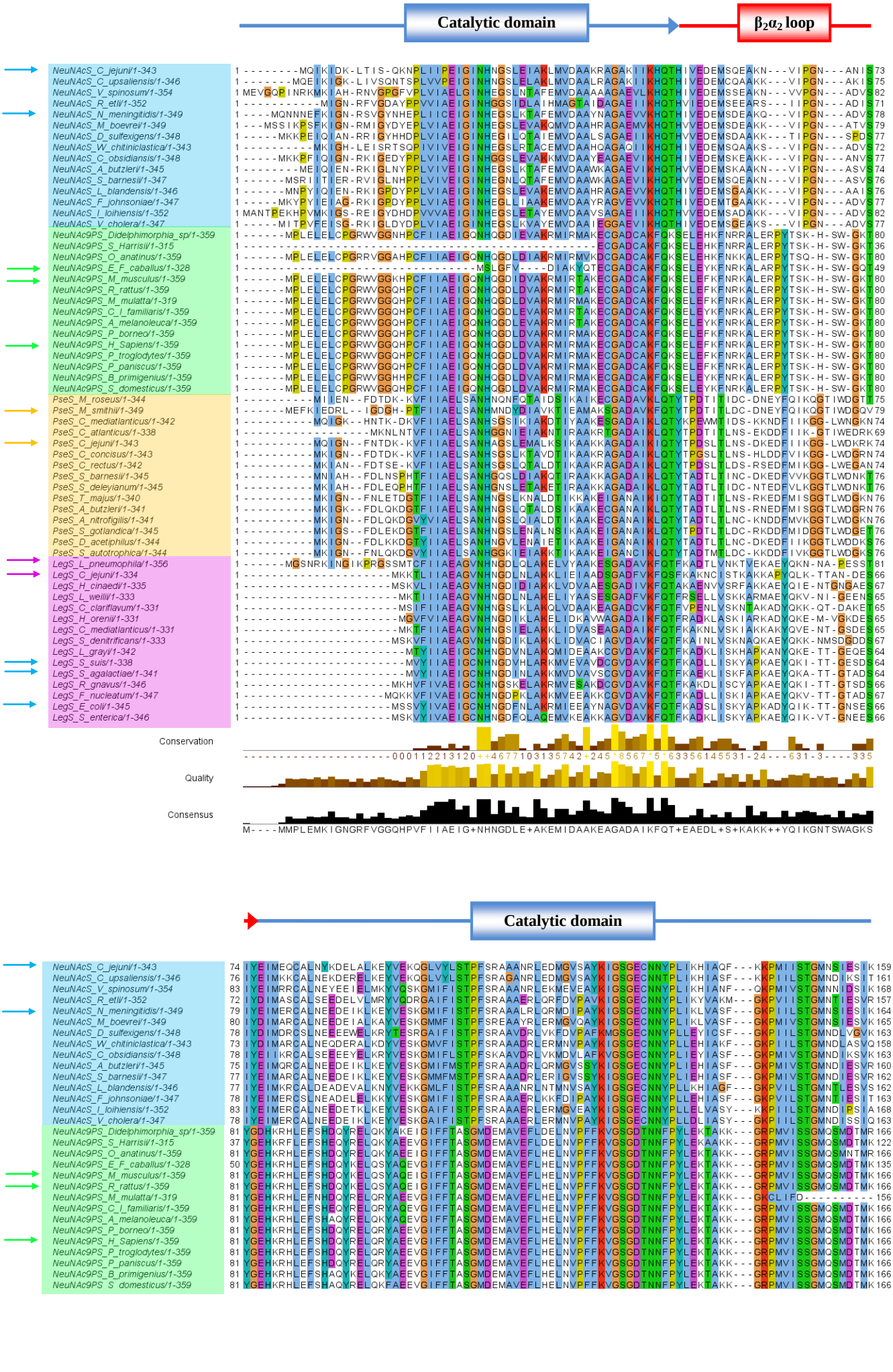
architectures that facilitate the recognition and catalysis of structurally variable sugar-substrates may assist with re-engineering such enzymes as sources of novel and complex sialic acids with useful applications.

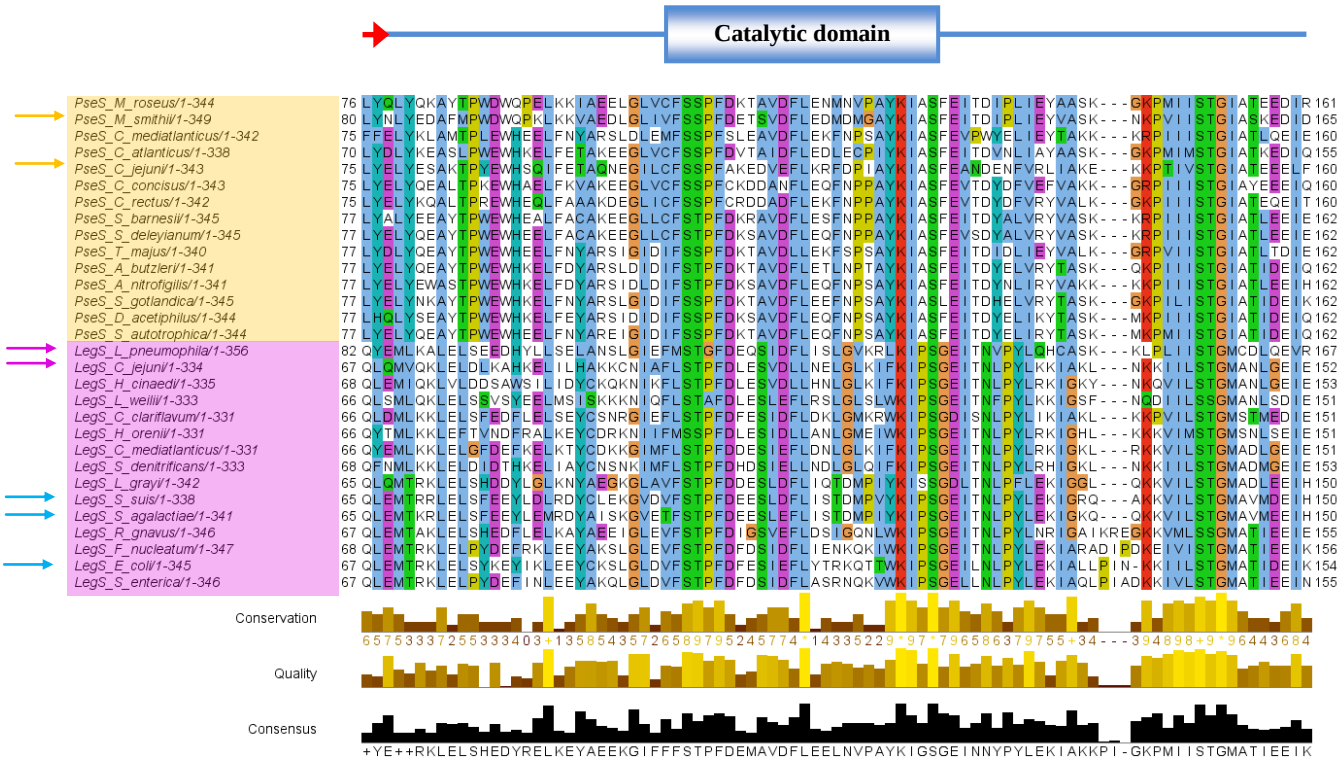
In the studies described in this chapter, sequence alignment of the aforementioned bacterial sialic acid synthase orthologues with other sialic acid synthase clades was used in an attempt to identify putative sugar binding motifs that may provide insight into how these enzymes recognise and bind substrates with variable stereochemistry and functional group substitution. Using the same methodology outlined in chapter three, two loop-swap chimeras were generated and characterised, in which the putative sugar-substrate binding loop ($\beta_2\alpha_2$ loop) sequence from *Cje*NeuNAcS was installed into the respective positions of both *Cje*PseS and *Cje*LegS in an attempt to confer (or improve) activity with the NeuNAcS substrate ManNAc. It should be noted that the generation of these chimeras was completed in parallel to those detailed in chapter three.

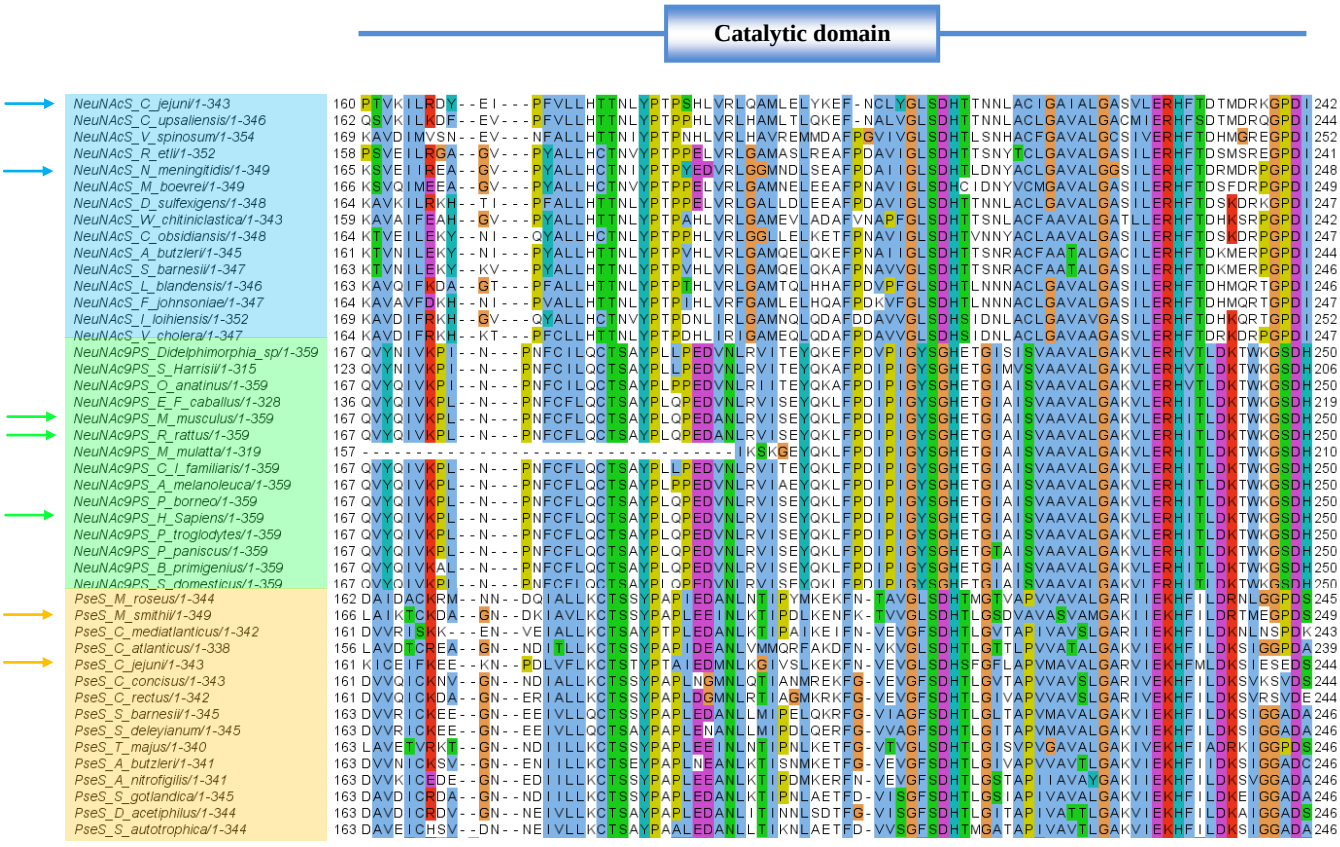
4.2 Sequence alignment of multiple sialic acid synthases clades

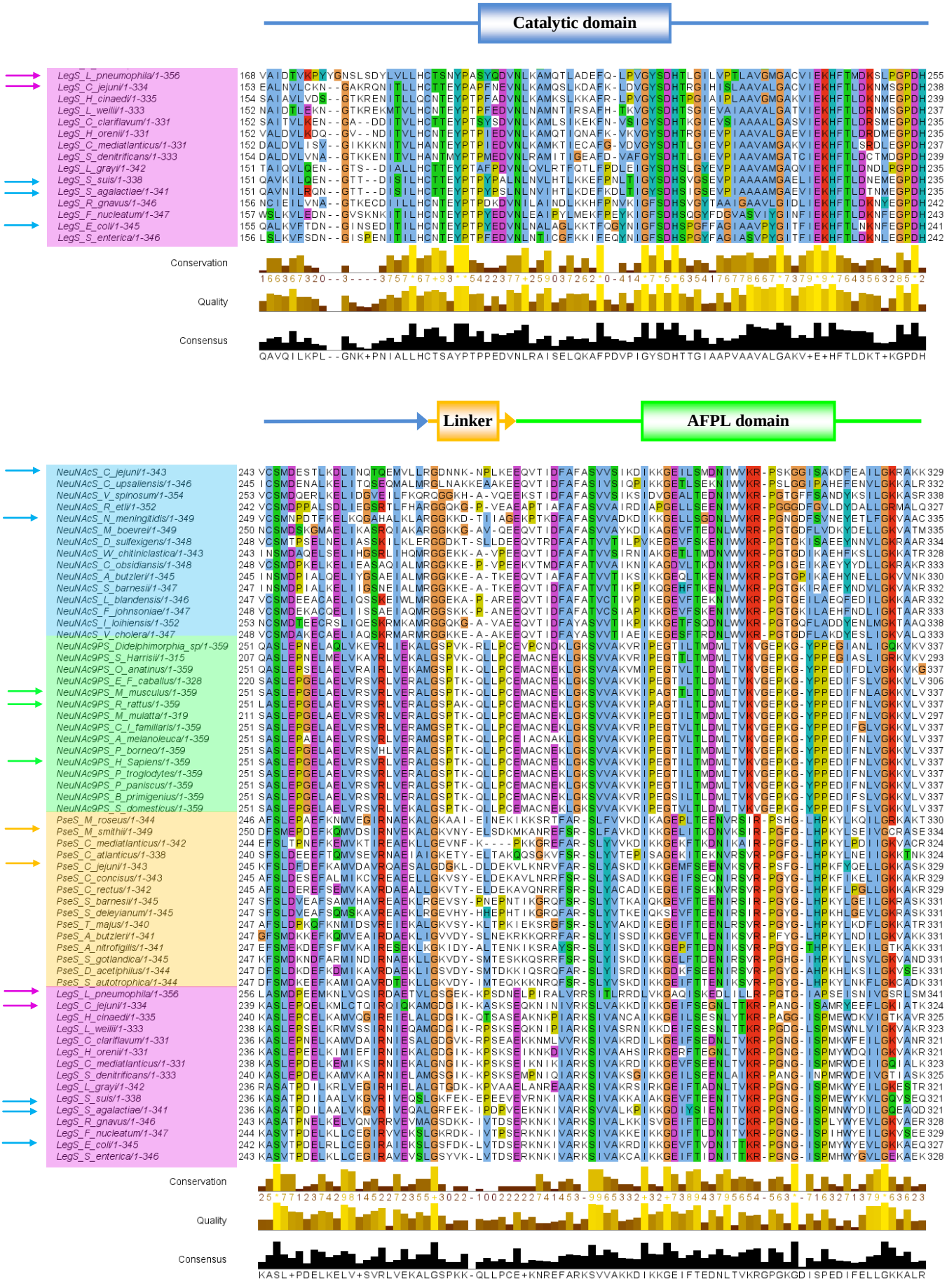
Dr. Joseph had previously generated a multiple sequence alignment of members of the sialic acid synthase family, including those assigned to NeuNAcS, PseS and LegS clades.[86] Additionally, a fourth clade, denoted by Dr. Joseph as 'Type IV sialic acid synthases' was included in the alignment.[86] This fourth clade encodes putative sialic acid synthases of currently unknown function. Unlike other members of the sialic acid synthase family, Type IV enzymes do not contain a peptide sequence corresponding to an AFPL domain, and thus are fundamentally contrasting to the other sialic acid synthases.

Interestingly, a number of sialic acid synthases previously characterised as having NeuNAcS activity, align within the putative LegS clade. Dr. Joseph suggested that these enzymes may possess dual-functionality, with the ability to utilise both the LegS precursor, and ManNAc in catalysis.[86] This is feasible as *Cje*LegS was shown to exhibit activity with both the LegS precursor, and ManNAc albeit very poor.[11, 86] Because of the commercial unavailability of the natural LegS substrate, and difficulty of producing it synthetically, most putative LegS enzymes are uncharacterised and denoted as LegS based on sequence comparisons alone. As such, the NeuNAcS enzymes that group within the LegS clade have not been confirmed as having LegS activity. Here I present a similar multiple sequence alignment, but have additionally included members of the NeuNAc 9-PS clade, and excluded 'Type IV' sequences given the nature of their substrates is unknown (Figure 4.1).


Catalytic domain
β2α2 loop
Conservation
Quality
Consensus
Catalytic domain


Catalytic domain
PseS_M_roseus/1-344
PseS_M_smithii/1-349
PseS_C_mediatlanticus/1-342
PseS_C_atlanticus/1-338
PseS_C_jejuni/1-343
PseS_C_concisus/1-343
PseS_C_rectus/1-342
PseS_S_barnesii/1-345
PseS_S_deleyianum/1-345
PseS_T_majus/1-340
PseS_A_butzleri/1-341
PseS_A_nitrofigilis/1-341
PseS_S_gotlandica/1-345
PseS_D_acetiphilus/1-344
PseS_S_autotrophica/1-344
LegS_L_pneumophila/1-356
LegS_C_jejuni/1-334
LegS_H_cinaedi/1-335
LegS_L_weilii/1-333
LegS_C_clariflavum/1-331
LegS_H_oreni/1-331
LegS_C_mediatlanticus/1-331
LegS_S_denitrificans/1-333
LegS_L_gryi/1-342
LegS_S_suis/1-338
LegS_S_agalactiae/1-341
LegS_R_gnavus/1-346
LegS_F_nucleatum/1-347
LegS_E_coli/1-345
LegS_S_enterica/1-346
Conservation
Quality
Consensus
+YE++RKLELSHEDYRELKEYAEEKGIFFFSTPFDEMAVDFLEELNVPAYKIGSGEINNYPYLEKIAKKPI-GKPMIISTGMATIEEIK


Catalytic domain
NeuNAcS_C_jejuni/1-343
NeuNAcS_C_upsaliensis/1-346
NeuNAcS_V_spinosum/1-354
NeuNAcS_R_etli/1-352
NeuNAcS_N_meningitidis/1-349
NeuNAcS_M_boevrei/1-349
NeuNAcS_D_sulfexigens/1-348
NeuNAcS_W_chitiniclastica/1-343
NeuNAcS_C_obsidiansis/1-348
NeuNAcS_A_butzleri/1-345
NeuNAcS_S_barnesii/1-347
NeuNAcS_L_blandensis/1-346
NeuNAcS_F_johnsoniae/1-347
NeuNAcS_I_loihiensis/1-352
NeuNAcS_V_cholera/1-347
NeuNAc9PS_Didelphimorphia_sp/1-359
NeuNAc9PS_S_Harrisii/1-315
NeuNAc9PS_O_anatinus/1-359
NeuNAc9PS_E_F_caballus/1-328
NeuNAc9PS_M_musculus/1-359
NeuNAc9PS_R_rattus/1-359
NeuNAc9PS_M_mulatta/1-319
NeuNAc9PS_C_l_familiaris/1-359
NeuNAc9PS_A_melanoleuca/1-359
NeuNAc9PS_P_borneo/1-359
NeuNAc9PS_H_Sapiens/1-359
NeuNAc9PS_P_troglodytes/1-359
NeuNAc9PS_P_paniscus/1-359
NeuNAc9PS_B_primigenius/1-359
NeuNAc9PS_S_domesticus/1-359
PseS_M_roseus/1-344
PseS_M_smithii/1-349
PseS_C_mediatlanticus/1-342
PseS_C_atlanticus/1-338
PseS_C_jejuni/1-343
PseS_C_concisus/1-343
PseS_C_rectus/1-342
PseS_S_barnesii/1-345
PseS_S_deleyianum/1-345
PseS_T_majus/1-340
PseS_A_butzleri/1-341
PseS_A_nitrofigilis/1-341
PseS_S_gotlandica/1-345
PseS_D_acetiphilus/1-344
PseS_S_autotrophica/1-344


Catalytic domain
Linker
AFPL domain
Conservation
Quality
Consensus

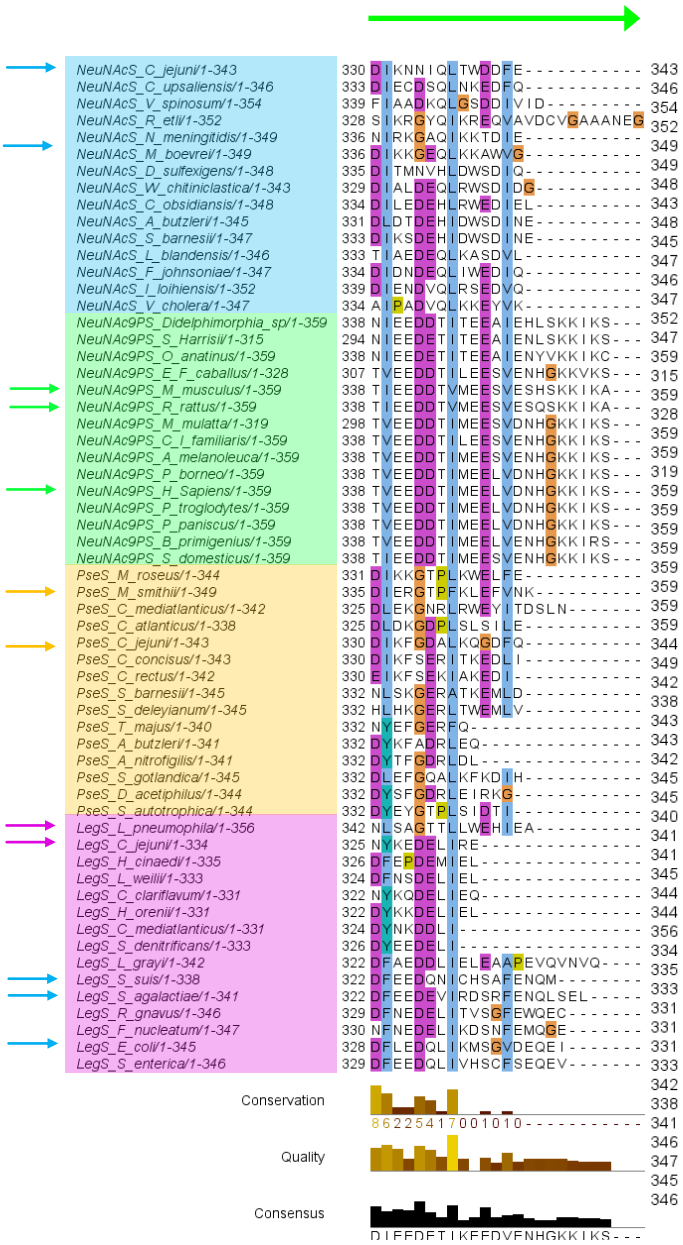


Figure 4.1. Sequence alignment of members of the sialic acid family showing four distinct clades. The first of these is that including bacterial NeuNAcS (blue). The second group includes mammalian NeuNAc 9-PS (green). A third group contains sequences most similar to that of enzymes previously characterised as PseS (orange). The fourth group contains sequences which align most closely with enzymes previously shown to have LegS activity (purple). Orthologues which have previously been functionally characterised are indicated by arrows. It should be noted that a number of bacterial enzymes characterised as NeuNAcS align with sequences of the putative LegS family (blue arrows). It is possible that these enzymes show bi-functional LegS and NeuNAcS activity. Domain annotation is based on the structure of *Nme*NeuNAcS (PDB: 1XUZ) and sequences are coloured according to the ClustalX colour scheme.

The two sequence elements that differ most notably between bacterial NeuNAcS and mammalian NeuNAc 9-PS, the $\beta_2\alpha_2$ loop and AFPL domain/linker, are also poorly conserved

between LegS and PseS clades. Again, this potentially implicates these regions of sequence in sugar-substrate binding. Dr. Joseph identified a conserved Ala-X-Tyr-Gln motif within the $\beta_2\alpha_2$ loop sequence of all examined LegS orthologues, where X is a variable residue. Homology modelling and substrate docking of *CjeLegS* did not predict any interactions of the $\beta_2\alpha_2$ loop residues with the sugar substrate, however given the poor sequence identity of the loop region with the homology model template sequence (*Nme*NeuNAcS), positioning of this loop is likely to be inaccurate.[86]

Conversely the fully conserved residues of the PseS $\beta_2\alpha_2$ loop are sparsely distributed across the entire loop, with no distinct motif formation.[86] As for LegS, a model of *Cje*PseS was also generated and docked with natural substrate *in silico*, however due to low template sequence identity, residues of the $\beta_2\alpha_2$ loop were again poorly modelled.[86]

Upon closer inspection of the LegS clade sequences, there appears to be further division into two smaller conserved groups. The LegS sequences chosen for alignment separate into two potential 'sub-clades' based on sequence alone (Figure 4.2). The first of these supposed sub-clades (sub-clade one) includes both of the enzymes characterised as having LegS activity (*Cje*LegS and *Lpn*LegS), while the second sub-clade (sub-clade two) contains three sequences which have been previously characterised as having NeuNAcS activity.[4, 55, 86] It is possible that sub-clade two represents a group of bi-functional LegS/NeuNAcS, while the other LegS sub-clade represents those with LegS only activity, however this is purely speculative. *Cje*LegS, which is known to have bi-functional LegS/NeuNAcS activity, however, groups within the first sub-clade.[86] Such assumptions of sub-clade separation are made under the proviso that the apparent disparity in sequence could be an artefact of the particular LegS sequences selected for alignment, or the limited number of sequences available for alignment. Further functional information relating to the members of these putative sub-clades is required to assist in delineation of the relationship between sequence and substrate specificity.

The significance of these putative sub-clades is speculative however there is considerable difference in sequence within, and between these groups. Sequence identity ranges from 44 - 75% within sub-clade one, and 59- 81% within sub-clade two. Between the two putative sub-clades, identity is lower, ranging between 39 and 58%.

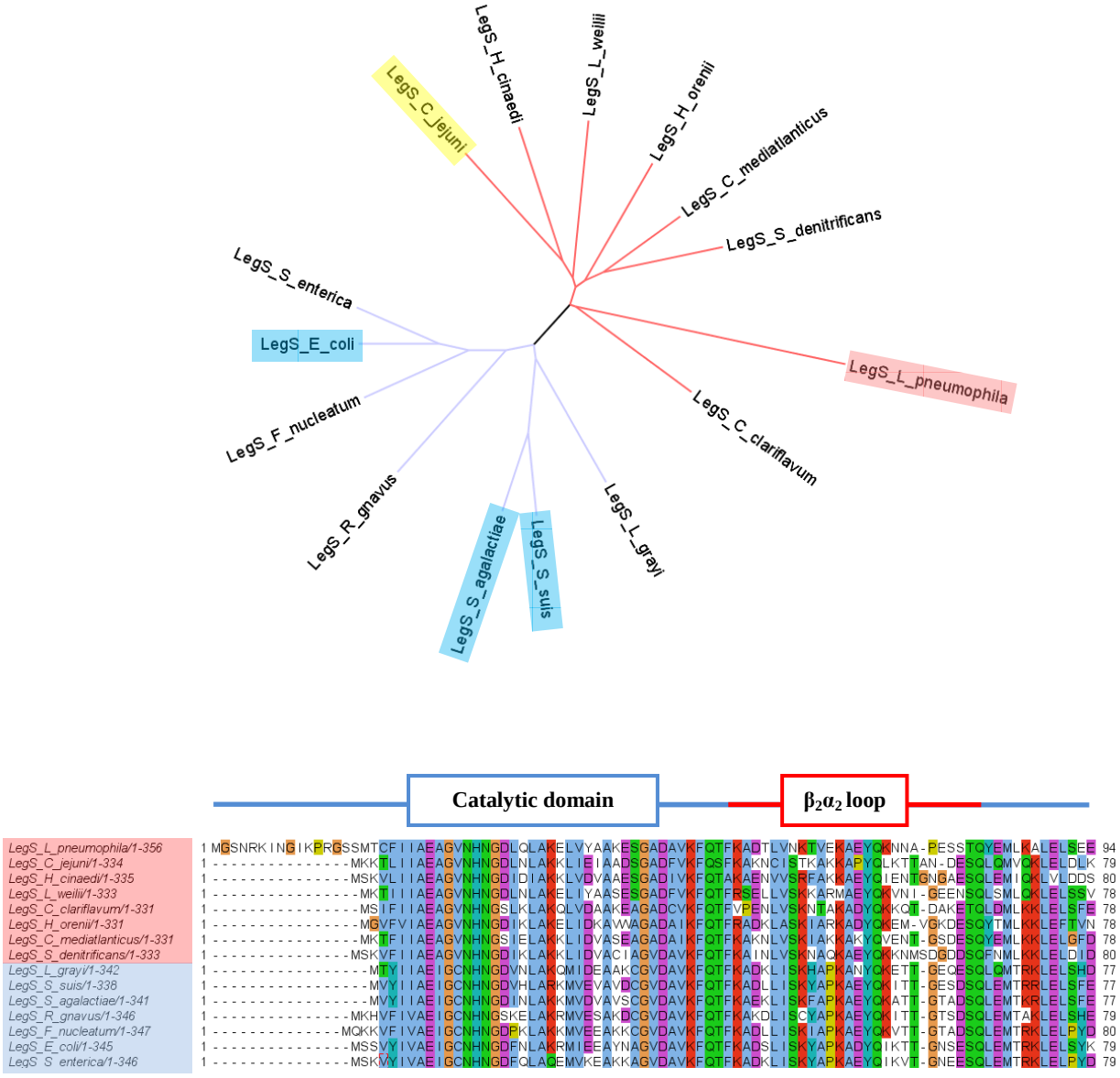


Figure 4.2. Phylogenetic tree (top) showing division of LegS sequences into postulated sub-clades one (red branches) and two (blue branches). Enzymes previously characterised as having LegS activity are highlighted red, and those known to have NeuNAcS activity are highlighted blue. *Cje*LegS which is characterised as having both LegS and NeuNAcS functionality is shown in yellow. Partial sequence alignment (bottom) of putative LegS sequences showing sub-clade 1 (red) and sub-clade 2 (blue) $\beta_2\alpha_2$ loop sequences.

Within the $\beta_2\alpha_2$ loop of putative sub-clade two, is a highly conserved Thr-Thr-Gly motif, which is otherwise poorly conserved in the putative LegS only sub-clade (Figure 4.2). Interestingly, *Cje*LegS (which is known to exhibit poor NeuNAcS activity) is the only sequence from sub-clade one to retain the threonine pair of this motif. Thr-61 (*C. jejuni*) of this motif is conserved as Asn-74 in NeuNAcS, a residue known to hydrogen bond to the C-6 hydroxyl of ManNAc.[32] Retention of a similar polar functionality at this position may allow

the enzymes of sub-clade two (and *Cje*LegS) to form a similar hydrogen bonding interaction with ManNAc. The denoted PseS sequences do not separate into obviously distinct sub-clades as observed for LegS, consistent with the fact that to date no PseS orthologues have been shown to exhibit bi-functional activity.

4.3 LegS and PseS $\beta_2\alpha_2$ loop-swap chimera design and generation

LegS from *C. jejuni* is characterised as having both a primary LegS function and secondary NeuNAcS function.[11, 86] This requires the ability to recognise and bind both the Leg precursor sugar 2,4-diacetamido-2,4,6-trideoxy-D-*manno*pyranose, and the NeuNAc precursor, ManNAc. The genome of *C. jejuni* also encodes a dedicated NeuNAcS (*Cj1141*), which has a specificity constant for ManNAc over a 1000-fold greater than that of bi-functional *Cje*LegS.[86] *Cje*PseS (*Cj1317*) is thought to be selective only for the sugar substrate 2,4-diacetamido-2,4,6-trideoxyaltrose. Lack of sequence homology across the $\beta_2\alpha_2$ loop between NeuNAcS, LegS and PseS sequences, suggests this region may play a role in selecting for and stabilising the respective sugar-substrates that these enzymes preferentially utilise in catalysis.

In order to probe the role of this loop in sugar-substrate selectivity of the aforementioned bacterial sialic acid synthases, two chimeric proteins were generated in which a block of 12 residues from the *Cje*NeuNAcS $\beta_2\alpha_2$ loop was engineered into the respective positions of *Cje*LegS and *Cje*PseS (Figure 4.3). The residues chosen for substitution are predicted to constitute the flexible portion of the *Cje*NeuNAcS $\beta_2\alpha_2$ loop based on sequence identity with *Nme*NeuNAcS. Additionally, the 12 residues chosen for substitution contain the conserved putative NeuNAcS sugar-substrate binding motif (-Pro-Gly-Asn-) identified by Dr. Joseph.[86]

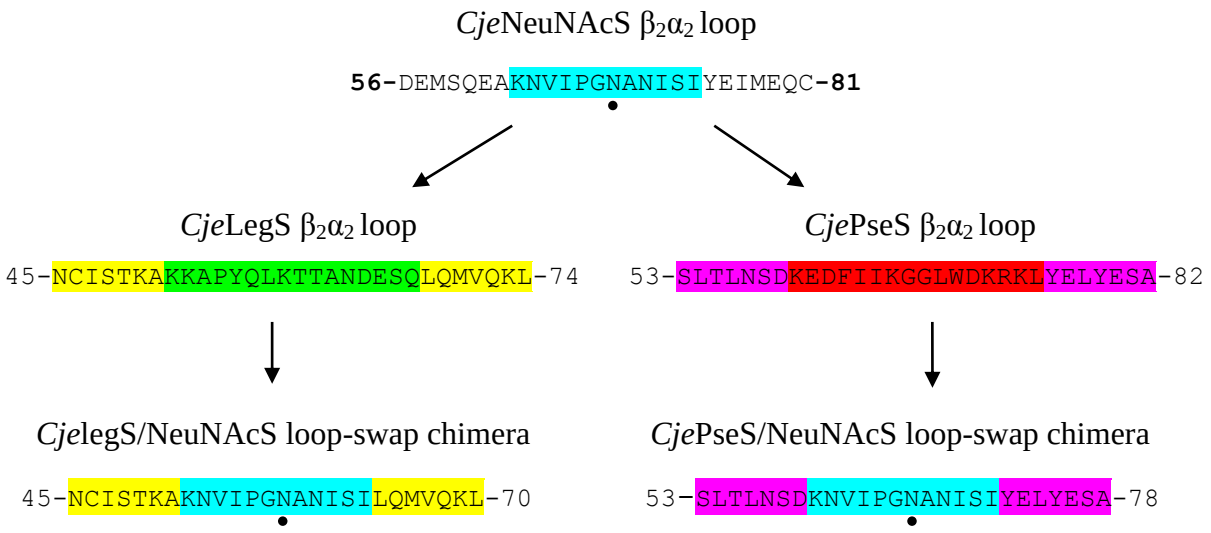


Figure 4.3. Residues from the $\beta_2\alpha_2$ loop of *Cje*NeuNAcS (blue) were substituted into the respective positions of *Cje*LegS and *Cje*PseS in an attempt to confer activity with the NeuNAcS substrate ManNAc. The ManNAc binding residue Asn-74 is indicated with a black circle.

Because the native sugar substrates of LegS and PseS were unavailable, the resulting chimeras were assayed for NeuNAcS activity to determine the effect of loop substitution. It was envisaged that installing the *Cje*NeuNAcS loop residues into *Cje*LegS would improve the already existing ability to utilise ManNAc. Conversely, *Cje*PseS has no bi-functional character and so it was hoped that by installing the NeuNAcS loop, a novel ability to utilise ManNAc would be conferred.

4.4 Cloning LegS and PseS loop-swap chimeras

Gene constructs encoding the aforementioned LegS and PseS loop-swap chimeras were generated using the same methodology as for the *Nme*/*Hsa* loop-swap chimera described in chapter three (Section 3.4). The genes encoding wild-type *Cje*LegS (Cj1327) and *Cje*PseS (Cj1317) had previously been cloned from genomic DNA (*C. jejuni* NCTC 11168) by Dr. Joseph.[86] Forward and reverse primers encoding the two halves of the *Cje*NeuNAcS loop residues to be inserted into the LegS and PseS scaffolds were designed (Table 6.4), and used to amplify pre and post-loop fragments which were subsequently extended using bridging primers. The resulting overlapping fragments were then extended using gene primers containing additional generic gateway sequences (Table 6.4). The completed linear constructs, with *CjeNeuNAcS* loop residues installed, were ligated into Gateway® destination vector pDEST^{TM}17 and sequence verified before proceeding to expression trials.

4.5 Expression and purification of LegS and PseS loop-swap chimeras

The pDEST^{TM}17 vectors bearing the loop-swap chimeras were transformed into *E. coli* BL21 (DE3) Star cells and *E. coli* BL21 (DE3) pBB540/pBB542 (Chaperone 3) cells for comparative expression testing. In order to test inducible expression of the target proteins, an induced culture (containing 0.5 mM IPTG) and non-induced culture of each cell line were grown overnight. Both chimeras were shown by SDS-PAGE to be over-expressed following induction with IPTG, and soluble in both cell lines. Again, chaperone 3 cells yielded the greatest quantity of soluble protein, and thus this cell line was chosen for upscaled expression and purification. The expressed proteins were purified using a similar protocol as described for wild-type *Hsa*NeuNAc 9-PS (Section 2.4) (Figure 4.4A).

Following purification, mass spectrometry confirmed both chimeras were of the correct molecular mass (Appendix B), and circular dichroism (CD) was used to ensure the retention of wild-type like secondary structure, and correct folding in solution (Figure 4.4B, 4.4C).

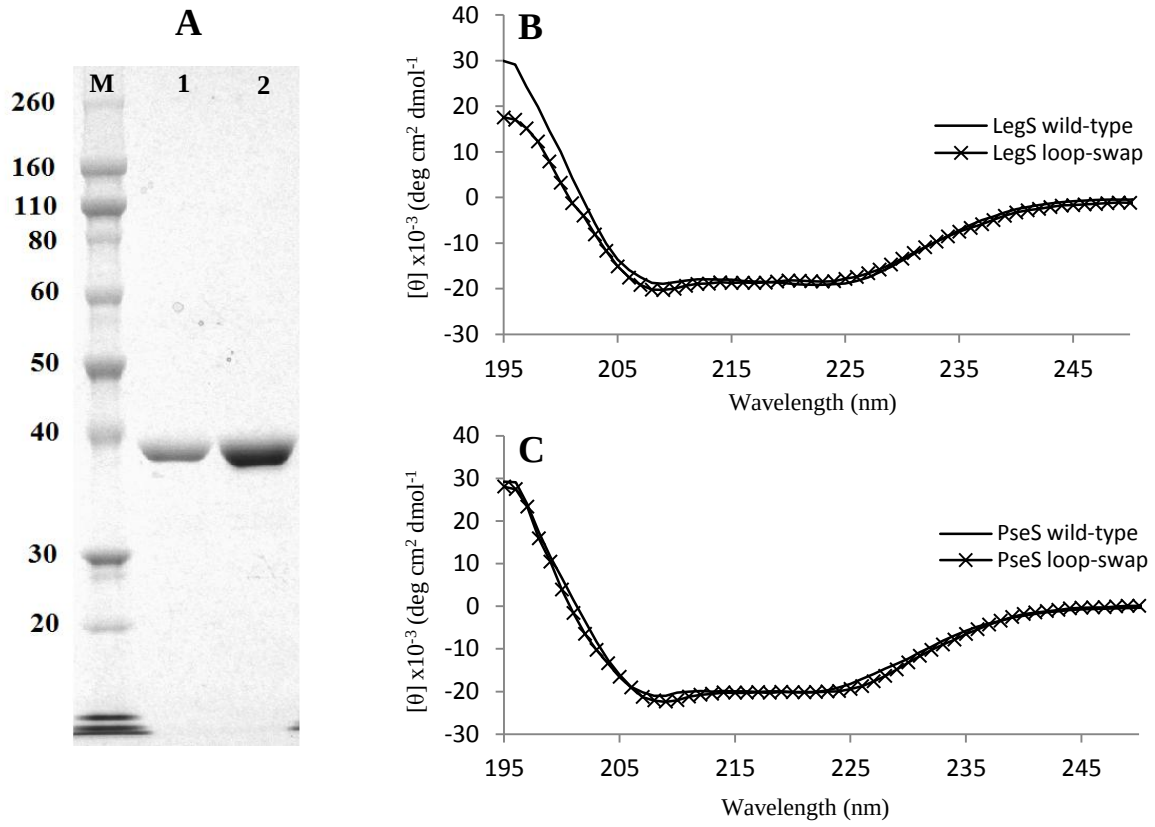


Figure 4.4. A) SDS-PAGE gel showing kDa molecular weight markers (M), purified PseS loop-swap chimera (1), and purified LegS loop-swap chimera (2). **B)** Comparison of CD spectra of wild-type *Cje*LegS and LegS loop-swap chimera. **C)** Comparison of CD spectra of wild-type *Cje*PseS and PseS loop-swap chimera.

4.6 Activity assessment of LegS and PseS loop-swap chimeras with ManNAc

Following purification, the two aforementioned bacterial sialic acid synthase loop-swap chimeras were assayed for activity with ManNAc using the continuous UV-Vis assay. The assay conditions used were the same as those established previously by Dr. Joseph for characterisation of wild-type *Cje*LegS and *Cje*PseS.[86] Under these conditions, no activity was detectable for either chimera. Furthermore, no NeuNAc formation was detected for either chimera after 24 hours, using the previously described NMR assay (Figure 4.5 & 4.6).

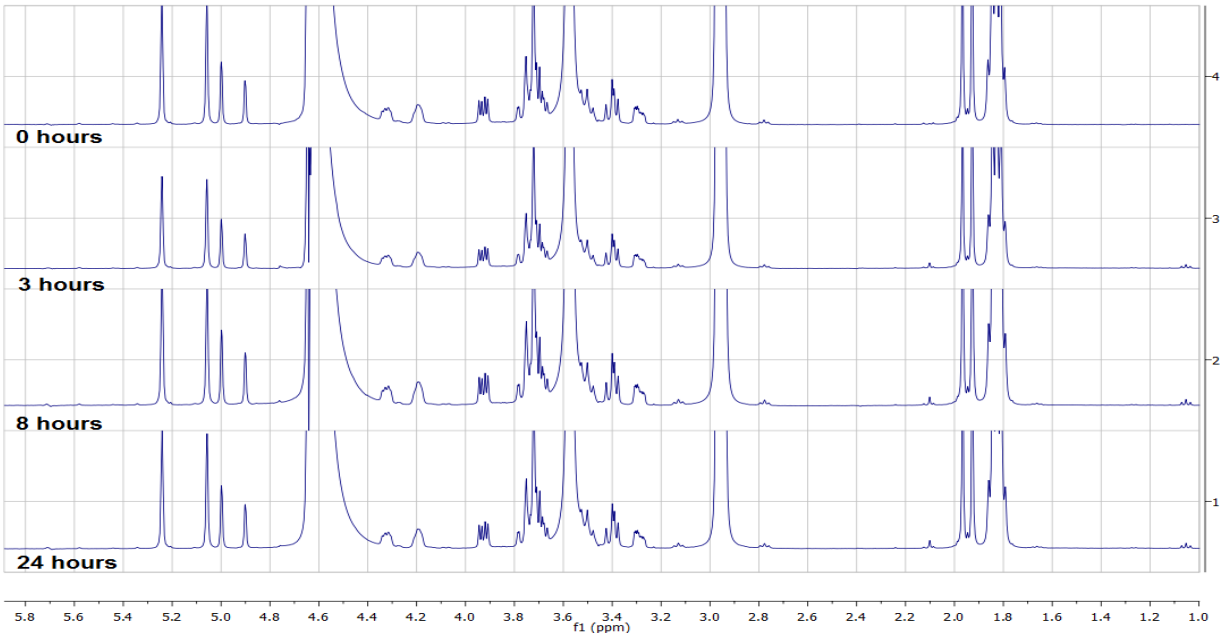


Figure 4.5. PRESAT NMR activity assay for LegS loop-swap chimera with 5 mM PEP and 5 mM ManNAc. No product formation is detectable after 24 hours.

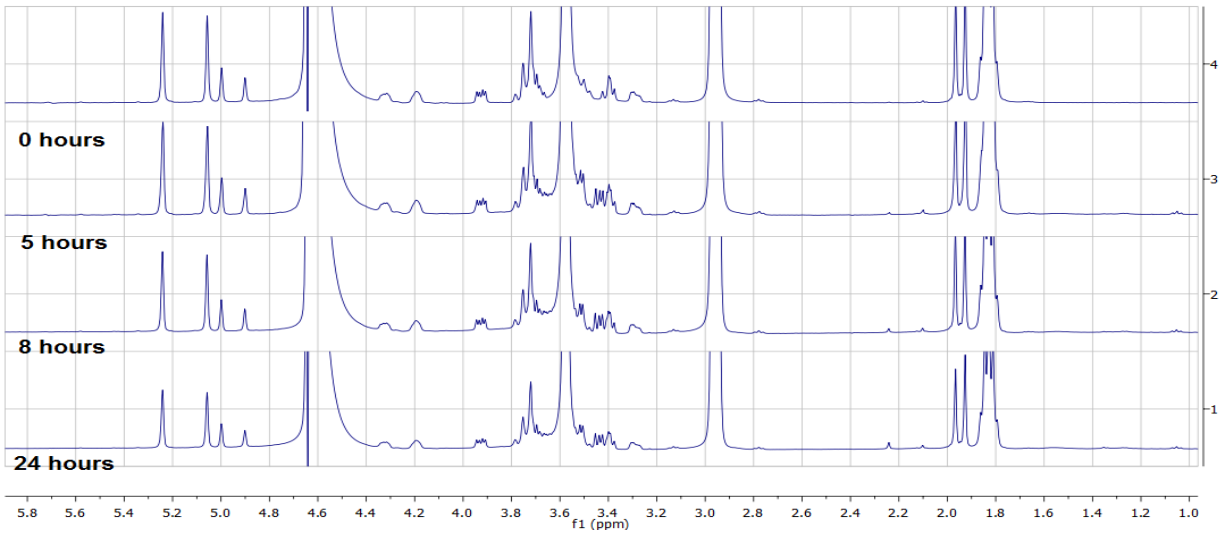


Figure 4.6. PRESAT NMR activity assay for PseS loop-swap chimera with 5 mM PEP and 5 mM ManNAc. No product formation is detectable after 24 hours.

4.7 PEP binding studies

The disappointing lack of activity observed for both chimeras in the presence of ManNAc and PEP warranted further investigation into the effects of loop substitution on substrate binding. To this end the PEP binding properties of the resulting chimeras were assessed by ITC (Figure 4.6).

The LegS loop-swap chimera was clearly shown to retain the ability to bind PEP, in fact the calculated dissociation constant for PEP was considerably lower than that of wild-type *Cje*LegS, suggesting improved PEP binding affinity. The enthalpic and entropic components of PEP binding appear to be similar to that of the wild-type enzyme (Table 4.1).

Wild-type *Cje*PseS was shown to have a similar K_d (PEP) to that of *Cje*LegS, and a similar overall change in free energy associated with PEP binding (Table 4.1). The specific thermodynamic parameters reported here for wild-type *Cje*PseS should however be viewed with scepticism as a number of erroneous data points had to be removed in order to fit the model. These anomalous data points most likely result from air trapped within the sample cell. When the same PEP titration was completed with the PseS loop-swap chimera, only a heat of dilution signal was observed, suggesting total abolition of PEP binding.

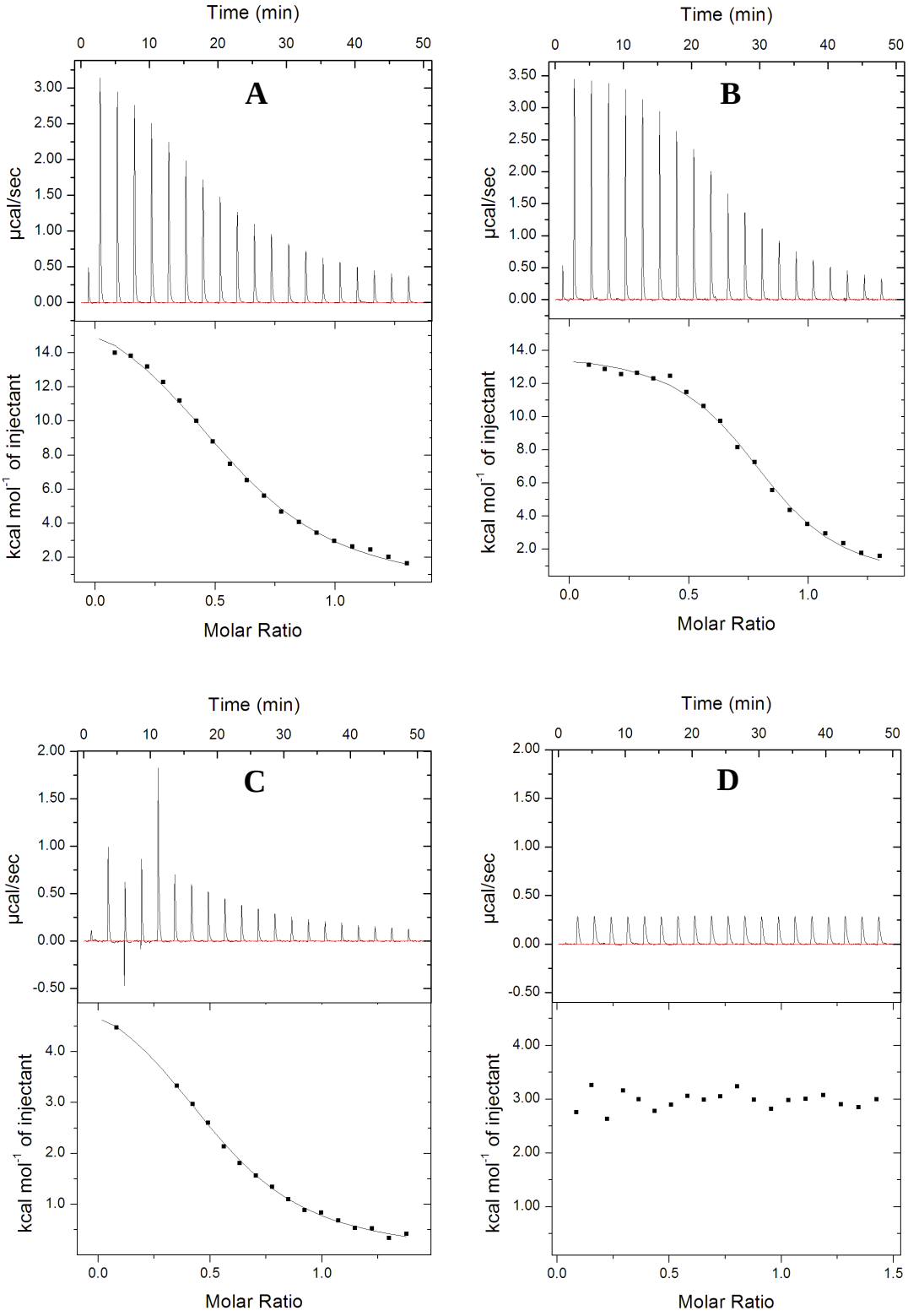


Figure 4.6. Raw injection peaks and integrated binding isotherms for titrations of **A)** 1 mM PEP into 150 µM wild-type *Cje*LegS, **B)** 1 mM PEP into 150 µM LegS loop-swap chimera, **C)** 1 mM PEP into 150 µM wild-type *Cje*PseS and **D)** 1 mM PEP into 150 µM PseS loop-swap chimera. All proteins were incubated with 1 mM $MnCl_2$ prior to titration with PEP.

Table 4.1. Calculated thermodynamic parameters of PEP binding

	Wild-type *Cje*LegS	LegS loop-swap chimera	Wild-type *Cje*PseS	PseS loop-swap chimera
K_d (PEP) µM	20 ± 1.5	6.1 ± 0.5	19 ± 1.0	-
ΔH cal/mol	18240 ± 430	13990 ± 180	5710 ±120	-
ΔS cal/mol/deg	83 ± 8.3	71 ± 7.1	41 ± 4.1	-

4.8 Structural studies

Dr. Joseph observed that wild-type *Cje*LegS and *Cje*PseS exist primarily as dimeric species in solution, under the conditions used for analytical gel filtration.[86] *Nme*NeuNAcS, which is known to adopt a functional domain-swapped homo-dimeric arrangement, shares high sequence identity with these enzymes and thus it is not surprising that they also exist as functional dimers. Analytical gel filtration was used to elucidate the native quaternary structures of the generated LegS and PseS loop-swap chimeras in solution. It is possible that disruption of native quaternary structure by introduction of *Cje*NeuNAcS loop residues is responsible for the inactivity of these chimeras (Figure 4.7).

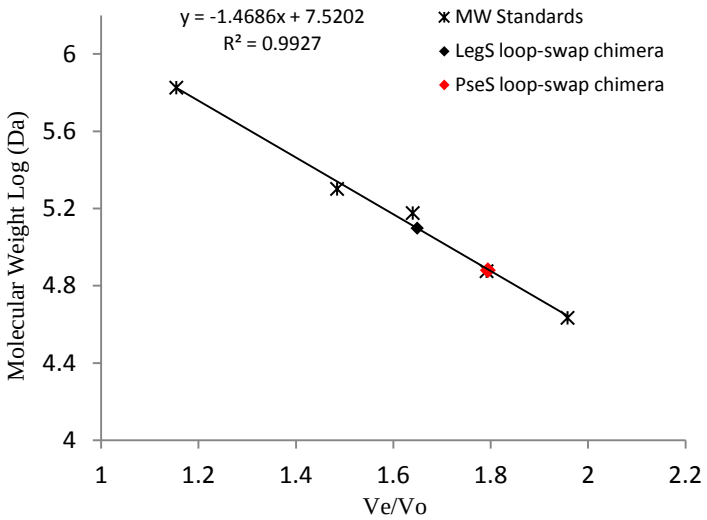


Figure 4.7. Semi-logarithmic plots showing elution volume of LegS and PseS loop-swap chimeras relative to molecular weight standards. Standards used include: thyroglobulin (669 kDa), beta amylase (200 kDa), alcohol dehydrogenase (150 kDa), ovalbumin (43 kDa) and conalbumin (75 kDa).

The PseS loop-swap chimera eluted from the column as a distinct, single peak at 13.6 mL. This elution volume gives an experimental molecular weight of approximately 75.9 kDa, which is in close agreement with the theoretical molecular weight of the dimeric species

(Table 4.2). The LegS loop-swap chimera also eluted as a single peak, however elution was much earlier than expected, at 12.5 mL. This elution volume corresponds to an approximate molecular weight of 125 kDa, which is more than three times the calculated monomeric mass. This observation suggests that the LegS loop-swap chimera predominantly adopts an oligomeric arrangement in solution with a molecular mass most consistent with that of a trimeric species.

Table 4.2. Experimental parameters determined for loop-swap chimeras by analytical gel filtration

	Elution volume (mL)	Elution vol/void vol	Experimental molecular mass (Da)	Theoretical dimeric mass (Da)	Theoretical trimeric mass (Da)
LegS loop-swap chimera	12.5	1.65	125300	73562	110343
PseS loop-swap chimera	13.6	1.79	75900	76425	114636

Small angle X-ray scattering (SAXS) data was obtained for both LegS and PseS loop-swap chimeras in the presence and absence of PEP (Figure 4.8). The scattering data obtained for LegS in the absence of PEP is of relatively poor quality with low signal to noise at high *q*. In the presence of PEP, scattering data is significantly improved. The scattering profile of the LegS loop-swap chimera in the presence of 1 mM PEP differs in shape from that of the unliganded enzyme considerably at q ~0.1, indicative of change in the overall shape of the protein (Figure 4.8A).[89] This may be attributed to a conformational change induced by PEP binding, as observed for *Nme*NeuNAcS.[77] The fact that the scattering profile becomes better defined upon addition of PEP provides further evidence for PEP induced structural stabilisation. Based on ITC data, it is established that this chimera retains wild-type like PEP binding affinity. The scattering data for both unliganded and PEP bound LegS loop-swap chimera fits very poorly with the theoretical scattering profile predicted for *Nme*NeuNAcS based on crystallographic coordinates (Figure 4.8A).

The scattering profile of the PseS loop-swap chimera is not significantly changed by the addition of PEP (Figure 4.8B). Assuming *Cje*PseS is subject to the same PEP induced conformational rearrangement observed for *Nme*NeuNAcS, the lack of change in scattering profile suggests that this chimera has lost the ability to bind PEP. This reinforces the lack of PEP binding observed by ITC. Unlike the LegS loop-swap chimera, scattering data obtained

for the PseS loop-swap chimera fits more closely with the theoretical scattering calculated for *Nme*NeuNAcS based on crystallographic coordinates. This may suggest an overall structural commonality with *Nme*NeuNAcS, and retention of the domain-swapped oligomeric conformation in solution.

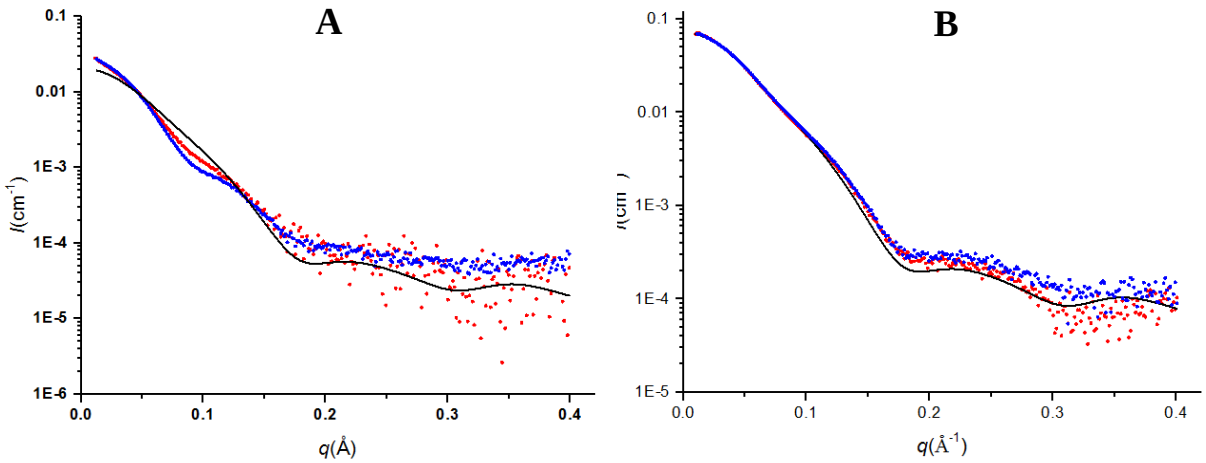


Figure 4.8. A) Small-angle X-ray scattering profiles of LegS loop-swap chimera in presence of 1 mM Mn^{2+} (red) and 1 mM Mn^{2+} + 1 mM PEP (blue) fit to the theoretical scattering profile predicted for *Nme*NeuNAcS (black line). **B)** SAXS profiles of PseS loop-swap chimera in presence of 1 mM Mn^{2+} (red) and 1 mM Mn^{2+} + 1 mM PEP (blue) fit to the theoretical scattering profile predicted for *Nme*NeuNAcS (black line).

Table 4.3. Experimental parameters obtained by small angle X-ray scattering (SAXS)

Structural parameters	LegS loop-swap chimera		PseS loop-swap chimera	
	1 mM $MnCl_2$	1mM $MnCl_2$ + PEP	1 mM $MnCl_2$	1mM $MnCl_2$ + PEP
$I(0)$ (cm^{-1}) [from P(r)]	0.03 ± 0.01	0.09 ± 0.01	0.07 ± 0.01	0.07 ± 0.01
R_g (Å) [from P(r)]	38.8 ± 1	39.8 ± 1	33.8 ± 1	33.2 ± 1
$I(0)$ (cm^{-1}) (from Guinier)	0.03 ± 0.01	0.09 ± 0.01	0.07 ± 0.01	0.07 ± 0.01
D_{max} (Å)	135 ± 7	134 ± 7	116 ± 6	111 ± 6
R_g (Å) (from Guinier)	39.5 ± 0.6	40.1 ± 0.7	33.9 ± 0.6	32.4 ± 0.2
Porod volume estimate ($Å^3$)	168300 ± 17000	208900 ± 21000	115300 ± 12000	114900 ± 11000
Dry dimeric volume calculated from sequence ($Å^3$)	88340	88340	91810	91810
Calculated monomeric molecular mass from sequence (Da)	36781	36781	38213	38213
Molecular mass from Porod volume (Da) (From MoW)	138000	171300	94530	94180

The calculated porod volumes derived from scattering data for the LegS loop-swap chimera vary considerably between the PEP bound and unliganded form (Table 4.3). In the absence of PEP, a porod volume of 168300 $Å^{-3}$ gives an approximate molecular mass of 138 kDa (calculated using MoW), which is most consistent with the theoretical mass of a tetrameric species. In the presence of PEP, this value increases to approximately 171 kDa which is more similar to that of a species with five subunits. There are several possible explanations for this observation. A combination of higher oligomeric state species may exist in solution, with PEP shifting the position of the oligomeric equilibrium to favour that of the higher molecular weight species. Size-exclusion chromatography used prior to analyses may not be able to fully resolve multiple species and therefore the final scattering profile may include contributions from more than one oligomeric state. Alternatively, the sample may be monodisperse for a single high molecular weight species but because of the poor-quality of scattering data, the derived porod volumes and molecular weights are not accurate enough to assign specific quaternary states to unliganded and PEP bound forms. What is apparent however is that introduction of the *Cje*NeuNAcS loop residues has somehow destabilised the native dimeric quaternary structure of *Cje*LegS, which may account for the loss of activity with ManNAc. Delineating the specific effect of loop substitution on the global solution structure of *Cje*LegS will require repetition of this SAXS experiment in order to obtain a higher quality data-set. Additionally, analytical ultracentrifugation (AUC) could be employed to further investigate the oligomeric arrangement of this chimera in solution.

Conversely, the structural parameters determined for the PseS loop-swap chimera are largely unchanged between PEP bound and unliganded forms. This provides further evidence for the absence of PEP binding. The calculated porod volume of the PseS loop-swap chimera gives an approximate molecular mass most consistent with a dimeric species (Table 4.3). Unlike the LegS loop-swap chimera, SAXS demonstrates that the PseS loop-swap chimera retains wild-type like quaternary structure, corroborating the aforementioned analytical gel filtration results.

4.9 Attempted crystallisation

The LegS and PseS loop-swap chimeras were screened for crystals using commercially available crystallisation conditions (Molecular Dimensions PACT, JCSG and Clear Strategy I & II). Protein was screened at a range of concentrations (2, 5 10 mg/mL) in the presence and absence of ligands, however crystals of diffraction quality were not obtained.

4.10 Discussion

Sequence analysis reveals a number of distinct sialic acid synthase clades exist across phyla. These sequence homology derived clades can be loosely annotated based on the function of known examples from within each clade, but such functional information remains sparse. Bacterial NeuNAcS and mammalian NeuNAc 9-PS are two distinct and obvious sialic acid synthase families, with a number of fully characterised enzymes from both. These families differ primarily by the sugar-substrates they utilise in catalysis. For NeuNAcS and NeuNAc 9-PS, the difference lies in the presence or absence of a phosphate moiety at C-6 of the hexose sugar-substrate. The bacterial sialic acid synthases, LegS and PseS also utilise variable sugar substrates in catalysis however selection is based upon substitution of functional groups and inversion of stereochemistry at multiple centres, rather than phosphorylation state (Figure 4.9).

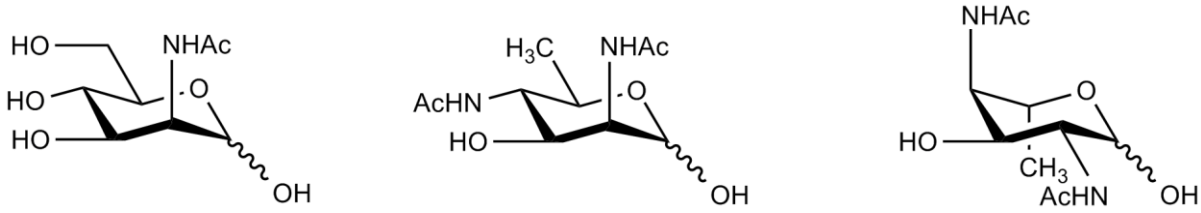


Figure 4.9. The structurally variable **s**ugar-substrates of the bacterial sialic acid synthases. The NeuNAcS sugar-substrate ManNAc (left), the LegS sugar-substrate 2,4-diacetamido-2,4,6-trideoxy-D-*manno*pyranose (centre) and the PseS sugar-substrate 2,4-diacetamido-2,4,6-trideoxyaltrose (right).

The genome of *Campylobacter jejuni* encodes all three of the aforementioned bacterial sialic acid synthases (NeuNAcS, LegS, PseS), suggesting a distinct and specific enzyme is required for the biosynthesis of each of the respective bacterial sialic acids.[11] This becomes convoluted by the fact that the product of the gene encoding LegS functionality in *C. jejuni* also exhibits poor NeuNAcS activity and thus can be considered bi-functional.[11, 86] The *C. jejuni* genome also contains a gene encoding a dedicated NeuNAcS enzyme, with much better catalytic properties than the aforementioned *Cje*LegS.[54] The minimal NeuNAcS activity observed for *Cje*LegS suggests that LegS may represent an evolutionary prototype, which would likely have been promiscuous with respect to its sugar-substrate.[86] This ancestral LegS would have subsequently become specialised for catalysis with specific sugar-substrates. Legionaminic and pseudaminic acid are directly linked to basic bacterial cellular functions, including maturation and functionalisation of flagellar proteins required for motility and

pathogenicity.[11, 44, 50] These roles are arguably more fundamental to survival and proliferation than the molecular mimicry role provided by NeuNAc and so it seems intuitive that the enzymes of legionaminic and pseudaminic acid biosynthesis should precede those of NeuNAc biosynthesis evolutionarily.

The additional advantage of *de-novo* NeuNAc biosynthesis for purposes of immune evasion would then provide the selection pressure required for specialisation of an ancestral LegS type enzyme toward that of modern day bacterial NeuNAcS. It is possible that enzymes currently characterised as NeuNAcS may exhibit residual LegS activity, which would provide evidence for divergence from a promiscuous LegS-like ancestor. Assaying for LegS activity with the appropriate sugar-substrate would be required to determine if this is indeed the case. Furthermore, assessment of bacterial viability following knockout of PseS, LegS and NeuNAcS genes from *C. jejuni* may provide some insight into the relative importance of these enzymes in conferring survival advantage to the species, and thus help confirm which clade represents the evolutionary prototype. The presence of archael sequences across the sialic acid synthase clades supports divergence from an ancestral precursor early in evolution,[90] as opposed to the otherwise proposed origin of bacterial sialic acid synthases through lateral gene transfer from mammals.[31]

Regardless of the origin of sugar-substrate specificity, sequence information would lead us to believe that the mechanism of sugar-substrate differentiation involves variation in sequence about the $\beta_2\alpha_2$ loop of the catalytic barrel of these enzymes. Poor sequence homology between the $\beta_2\alpha_2$ loops of the alternate bacterial sialic acid synthases, but high conservation within each clade suggests a functional adaptation involving this loop. Given that these enzymes differ only by the sugar-substrates they utilise, one can assume the involvement of the loop region in discrimination between substrates. As described in chapter three, introduction of the human NeuNAc 9-PS $\beta_2\alpha_2$ loop into the *Nme*NeuNAcS scaffold was unsuccessful at conferring novel activity with the mammalian substrate. It is expected that failure to achieve this was in part due to relatively low sequence homology between *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS (29%). In this chapter I describe the generation of an additional two loop-swap chimeras in which the $\beta_2\alpha_2$ loop from *Cje*NeuNAcS was installed into the respective positions of *Cje*LegS and *Cje*PseS. These bacterial sialic acid synthases share considerably higher sequence identity with one another, than bacterial NeuNAcS shares with mammalian NeuNAc 9-PS. *Cje*NeuNAcS and *Cje*LegS are 36.7 % identical at the amino acid level, while *Cje*NeuNAcS and *Cje*PseS share 32.6% identity (Appendix A). It was

envisaged that this increased sequence identity between loop 'donor' and loop 'acceptor' enzymes would increase the likelihood of successful transfer of improved or novel activity.

Both LegS and PseS loop-swap chimeras were successfully cloned, expressed and purified to homogeneity as soluble protein. These novel proteins were shown to be correctly folded in solution, and maintained wild-type like secondary structure composition. Unfortunately, both were inactive with respect to the sugar-substrate ManNAc. Further analysis revealed the PseS loop-swap chimera had completely compromised PEP binding ability, with no binding event detectable by ITC or SAXS. The lack of a crystal structure or robust model of the wild-type enzyme makes it extremely difficult to delineate potential interactions between the $\beta_2\alpha_2$ loop residues and PEP binding residues. It is however apparent that substitution of the PseS loop residues with those of *Cje*NeuNAcS has had a drastic effect on PEP binding. Because this loop-swap chimera maintains a wild-type like quaternary arrangement in solution, loss of PEP binding is assumed to be the primary cause of inactivity.

Interestingly, the LegS loop-swap chimera retains full (if not improved) PEP binding capacity and thus represents the first of the chimeras described in this thesis to do so. This may be due to the relatively high sequence identity between *Cje*LegS and *Cje*NeuNAcS (36.7%). The substituted *Cje*NeuNAcS loop residues are obviously capable of compensating the loss of the wild-type residues with respect to whatever role they play in PEP binding. Despite maintenance of PEP affinity, the LegS loop-swap chimera has a significantly destabilised oligomeric state. It is unknown how substitution of the designated loop residues could result in such alteration of the quaternary structure however a subtle change in the physico-chemical properties of the protein surface may be sufficient to elicit this effect. It should be noted that the *Cje*NeuNAcS loop sequence installed in the chimera is four residues shorter than the substituted *Cje*LegS loop sequence. Truncation of this loop by four residues may have altered the conformation of adjacent secondary structural elements, thereby disrupting contacts between adjoining subunits. Loss of the functional dimeric conformation may be responsible for the resulting loss of activity with ManNAc.

Although inactive with respect to ManNAc, the LegS loop-swap chimera retains PEP binding affinity and so may provide a good scaffold for the continued pursuit of improved NeuNAcS activity. The next obvious step is to determine if this chimera is able to bind the sugar-substrate. This may be accomplished through ITC titrations with the non-reactive ManNAc analogue rManNAc. If this chimera is indeed able to bind rManNAc, it is likely that key

catalytic residues are displaced within the active site. Given what we know about the importance of active residues interacting across dimer interfaces in both *Nme*NeuNAcS and *Hsa*NeuNAc 9-PS, it is possible that disruption of quaternary structure here has yielded a catalytically incompetent active site.[86] The AFPL domain and linker region are known to not only contribute catalytic residues to the active site of the adjacent monomer, but also provide considerable stabilisation of dimeric quaternary structure through inter-domain interactions as shown for *Nme*NeuNAcS.[77] Perturbation of inter-domain interactions through substitution of the $\beta_2\alpha_2$ loop sequence may be responsible for the disruption of quaternary structure, and therefore loss of catalysis. Access to high-resolution crystal structures of these enzymes would provide the molecular insight required to delineate the sugar-substrate binding modes, and identify the cause of inactivity following substitution of $\beta_2\alpha_2$ loop-sequences. Crystallisation of these proteins should be prioritised in future work.

To conclude, variation observed in the $\beta_2\alpha_2$ loop sequence between the various bacterial sialic acid synthases is likely the result of evolutionary specialisation from a once promiscuous ancestral enzyme. Interchanging the central portion of the $\beta_2\alpha_2$ loop between these enzymes failed to interchange their sugar-substrate specificity due to unforeseen and confounding implications with PEP binding and quaternary structure destabilisation. Artificially recreating the variation in sugar-substrate specificity achieved by evolution over hundreds of millions of years is apparently no trivial task.

Chapter 5: Discussion and concluding remarks

"Sialic acids are not only the most interesting molecules in the world, but also the most important".[91]

This claim, while obviously subjective, is certainly arguable in the context of mammalian sialic acids. These compounds are expressed by almost every mammalian cell and tissue type, at all stages of development, and the range of functions they provide is not only extremely diverse, but also absolutely fundamental to cellular function.[1, 4, 92] It is of no surprise that the biosynthesis of the most common mammalian sialic acid NeuNAc has evolved to become a highly efficient process.

This efficiency is achieved in part, through utilisation of the phosphorylated sugar-substrate ManNAc 6-P by NeuNAc 9-PS, the central enzyme of NeuNAc biosynthesis in mammals. The analogous bacterial enzyme alternatively utilises the non-phosphorylated sugar ManNAc in catalysis. Chapter two of this thesis details the characterisation of human NeuNAc 9-PS, (*Hsa*NeuNAc 9-PS) and exploration of its active site architecture using homology modelling and directed mutagenesis. This work has exposed a number of key features of the human enzyme which allow it to efficiently bind its phosphorylated sugar-substrate, including a Thr-Lys phosphate binding pair which was identified as potentially important for stabilisation of ManNAc 6-P. Other related PEP-utilising aldolases including DAH7PS employ a similar motif for binding phosphorylated sugar substrates.[67] What is particularly interesting about *Hsa*NeuNAc 9-PS however, is that the lysine residue of the identified pair is not adjacent to the threonine in sequence, but instead is recruited from the opposite chain of the functional dimeric enzyme and extends across the dimer interface to fulfil its role. This finding rationalises the domain-swapped homo-dimeric arrangement that *Hsa*NeuNAc 9-PS was shown to form in solution. Furthermore, an abundance of positively charged residues across the putative sugar-substrate binding $\beta_2\alpha_2$ loop possibly acts in combination with specific phosphate binding residues to provide an electrostatic environment which favours binding and stabilisation of polyanionic ManNAc 6-P.

Adoption of a phosphorylated sugar-substrate with an inherently higher binding affinity confers considerable catalytic efficiency to the mammalian enzyme. Such advantage is evident in the vastly improved specificity constant of the human enzyme for its sugar substrate, relative to that of the bacterial enzyme for unphosphorylated ManNAc.[32]

The homology model generated for *Hsa*NeuNAc 9-PS, and interactions predicted by induced-fit substrate docking, were partially corroborated through mutagenesis of predicted sugar-substrate binding residues. Delineation of the precise sugar-substrate binding mode however will require access to a high-resolution crystal structure. While crystals of diffraction quality were not obtained during this study, a number of positive crystallisation leads were identified and provide a good starting point for the pursuit of a *Hsa*NeuNAc 9-PS crystal structure. If achieved, this would represent the first structure of a NeuNAc 9-PS, and the second only structure of any sialic acid synthase.

Certain pathogenic bacteria also express NeuNAc on their own cell surfaces, allowing immune system evasion by mimicking the surface physiology of their mammalian host's cells.[7] As previously mentioned, bacterial and mammalian sialic acid synthases differ primarily in the nature of the sugar-substrates they utilise in catalysis.[4] Bacterial NeuNAcS catalyses the condensation of PEP with ManNAc, whereas mammalian NeuNAc 9-PS catalyses the condensation of PEP with the phosphorylated derivative ManNAc 6-P. Interestingly, these enzymes are entirely exclusive for their respective sugar-substrates. Bioinformatic analysis of bacterial and mammalian sialic acid synthases reveals the sequences corresponding to the $\beta_2\alpha_2$ loop and AFPL domain/linker regions are most variable between these two clades, but are highly conserved within each clade.[86] This alludes to an evolutionary divergence of these sequences to facilitate alternate functions. The $\beta_2\alpha_2$ loop and AFPL domains are known to contribute sugar-substrate binding residues in both *Nme*NeuNAcS and *Hs*aNeuNAc 9-PS, and thus peptide sequence variation across these regions is likely indicative of involvement in selection for and stabilisation of their respective sugar-substrates.[32] Unlike the $\beta_2\alpha_2$ loop of mammalian NeuNAc 9-PS, which is rich in conserved cationic residues, the bacterial $\beta_2\alpha_2$ loop does not possess such positive charge character. This can be rationalised in that the bacterial sugar-substrate ManNAc lacks the C-6 phosphate group, and is neutral at physiological pH, meaning electrostatic stabilisation is unnecessary. Variance in electrostatic properties across the sugar-substrate binding site of mammalian and bacterial enzymes may reflect a mechanism of selecting for phosphorylated and unphosphorylated forms of the sugar-substrate.

While the $\beta_2\alpha_2$ loop and AFPL domains are likely to be involved in discriminating between bacterial and mammalian sugar-substrates, the sugar-substrate specificity of *Nme*NeuNAcS could not be altered to that of the human enzyme by simply substituting these regions of

interest. Low sequence identity between the two enzymes makes this a difficult task, with unforeseen disruption to PEP binding and quaternary structure hindering success.

In bacterial systems, NeuNAc is also utilised in catabolic pathways as a source of carbon, nitrogen and energy (Figure 3.22).[7, 88] The first step of NeuNAc catabolism involves lyase cleavage of NeuNAc to pyruvate and ManNAc. ManNAc is subsequently phosphorylated to ManNAc 6-P which eventually enters glycolysis as fructose 6-phosphate.[7] As bacterial NeuNAcS is specific only for unphosphorylated ManNAc, phosphorylation at C-6 would effectively commit ManNAc to the catabolic pathway, preventing futile return to the NeuNAc pool.

The absolute specificity for the unphosphorylated sugar-substrate also makes bacterial NeuNAcS a promising target for inhibition with novel antibiotic compounds. Because the bacterial and mammalian enzymes are specific for their respective sugar-substrates, it is unlikely that sugar-substrate or transition-state analogues designed to bind bacterial NeuNAcS, would have any inhibitory effect on the mammalian enzyme. Inhibition of bacterial NeuNAcS may provide a novel approach to the early stage treatment of infections otherwise resistant to common antibiotics. Disabling the *de-novo* NeuNAc biosynthesis pathway in pathogenic bacteria would effectively destroy the pathogen's molecular disguise, and expose them to the host's immune system for clearance.

In addition to NeuNAc, several species of pathogenic bacteria synthesise a number of unique sialic acids not observed in the animal kingdom.[10-11, 48] Legionaminic and pseudaminic acid are two such 'bacterial sialic acids', both of which are required for the glycosylation and function of cell surface structures including flagella, in a number of motile pathogenic bacteria.[44, 48] Involvement in the sialylation of flagellar proteins directly links these sugar compounds to the motility and virulence of these species.[11, 47] Legionaminic and pseudaminic acid are biosynthesised by condensation of PEP with 2,4-diacetamido-2,4,6-trideoxy-D-*manno*pyranose and 2,4-diacetamido-2,4,6-trideoxyaltrose respectively. These sugar-substrates differ from one another by inversion of stereochemistry, and from ManNAc by variation in the nature and stereochemistry of the hexose-ring substituents (Figure 4.9). The aforementioned condensation reactions are catalysed by members of the evolutionarily related sialic acid synthase family. How these enzymes selectively recognise and bind the appropriate hexose sugar-substrate *in vivo* remains unknown, however multiple sequence alignment again suggests involvement of the $\beta_2\alpha_2$ loop and AFPL domain/linker. It seems

likely that these regions, which are poorly conserved between the different sialic acid synthase clades, became specialised through divergent evolution from a promiscuous ancestral enzyme capable of utilising multiple sugar substrates in catalysis.[90] Legionaminic acid synthase (LegS) may represent an extant relic of this ancestral enzyme as *Cje*LegS is bi-functional in nature, and exhibits low levels of activity with ManNAc.[86] A straight forward substitution of the $\beta_2\alpha_2$ loop residues of *Cje*LegS and *Cje*PseS with those from *Cje*NeuNAcS failed to improve activity with ManNAc, suggesting sugar-substrate selectivity results from a far more complex set of interactions between the $\beta_2\alpha_2$ loop and other structural elements. Solving the crystal structures of these bacterial sialic synthases will allow identification of the specific sugar-substrate binding residues, and will aid in tracking the evolutionary changes which have allowed these enzymes to become specific for their respective substrates. Delineation of binding modes and key active site contacts may additionally provide the basis for design of bacterial sialic acid synthase inhibitors as novel antibiotic compounds. The absence of these enzymes in mammals and their critical role in motility and pathogenicity makes LegS and PseS ideal candidates for drug-targeting.

It is remarkable to think that the modern day sialic acid synthases are capable of recognising and binding the correct hexose sugar-substrates *in vivo*, considering the plethora of structurally similar compounds present in the cellular environment. This ability to discriminate between subtle structural differences in sugar-substrate makes these central enzymes of sialic acid biosynthesis a paradigm of molecular recognition. By further understanding the ways in which this specificity is achieved on a molecular level, we will begin to expose ways in which these enzymes can be exploited toward therapeutic or biosynthetic advantage.

Chapter 6: Materials and methods

6.1 Computational tools

6.1.1 Protein structure visualisation

Protein structures were visualised, and figures created, using PyMOL (version 1.5, Schrödinger LLC).[93]

6.1.2 Sequence alignments

Pairwise and multiple sequence alignments were generated using the ClustalOmega server.[83-84] Alignment outputs were visualised using JalView.[94]

6.1.3 Phylogenetic trees

The multiple alignment outputs from ClustalOmega were entered into Figtree for generation and visualisation of phylograms.[95]

6.1.4 Homology modelling and induced-fit substrate docking of HsaNeuNAc 9-PS

A model of *Hsa*NeuNAc-9-PS was built by homology modelling using Prime.[96-98] The model was constructed using the comparative method with the crystal structure of *Nme*NeuNAcS as the template (PDB ID: 1XUZ). Both monomers were built at the same time to generate the dimer structure. The ligands bound in the active site of *Nme*NeuNAcS (PEP and rManNAc) were retained in the homology model of *Hsa*NeuNAc-9-PS. The dimer was then minimised with Prime, to optimise bond length and bond angles, and to remove steric clashes.

The structure of the ligand ManNAc 6-P was built and prepared in Maestro and LigPrep respectively.[99-100] The modelling of ManNAc 6-P to the active site of *Hsa*NeuNAc 9-PS was conducted with the Induced-Fit Docking protocol from Schrodinger Suite 2012.[101-104] Since the ManNAc binding loop in the homology model shows different conformations in each monomer, induced-fit docking of ManNAc 6-P was conducted in each of the monomers. The centre of the grid was defined as centroid of the workspace ligand, i.e. the ManNAc molecule bound in the homology model, since ManNAc 6-P was expected to bind in the same pocket in the active site as ManNAc in *Nme*NeuNAcS. For the initial docking, the Van der Waals radii of the atoms of the ManNAc 6-P molecule and the enzyme homology model were scaled by a factor of 0.5. The 20 best solutions of the initial docking were kept. Residues 70-78 (on the ManNAc binding loop) as well as all residues on the homology model within a 5 Å

distance of the respective docked pose were refined. The ligands were re-docked to the top 20 newly generated structures if the energy was within 30 kcal/mol of that of the best pose.

6.2 General laboratory methods

6.2.1 Water

All water used in these experiments was purified using a Millipore Milli-Q system. Water for all molecular biology applications was autoclave sterilised prior to use.

6.2.2 Buffer preparation

All buffers were pH adjusted using a Mettler Toledo SevenCompact^{TM} 220 pH probe which was calibrated daily prior to use. pH adjustments were made using 1/10 M HCl and 1/10 M NaOH at the appropriate temperature. All buffers were filtered (0.2 µm) under vacuum and refrigerated until required.

6.2.3 Removal of metal ions

Divalent metal ions were removed from buffers by treatment with Chelex 100 Resin (Bio-Rad). 5 g of resin was added per 100 mL solution and stirred for 1 hour at room temperature. Chelex-treated buffers were then re-pH adjusted prior to removal of chelex resin by filtration (0.2 µm) under vacuum.

6.2.4 Enzyme substrate preparation

Substrates (PEP, ManNAc, ManNAc 6-P, Man 6-P) used for all kinetic assays were prepared as 10 or 100 mM stock solutions in the appropriate assay buffer, and filtered prior to use. Because PEP is highly acidic, the stock solution was pH adjusted before use. Accurate substrate concentrations were determined prior to kinetic assays, by measuring (in triplicate) the change in absorbance at 232 nm following enzymatic catalysis of the limiting substrate. Wild-type *Nme*NeuNAcS was used for determination of PEP and ManNAc concentration, and wild-type *Hsa*NeuNAc 9-PS for ManNAc 6-P and Man 6-P. A control sample lacking substrates was used to measure the change in absorbance upon enzyme addition. The total change in absorbance (ΔA) upon completion of the enzyme catalysed reaction [(starting absorbance + enzyme absorbance) - (final absorbance)] could be used to calculate the concentration of the limiting substrate with the Beer-Lambert equation using the extinction coefficient of PEP ($\varepsilon = 2.8 \times 10^3\ M^{-1}\ cm^{-1}$).

6.2.5 Antibiotics

All proteins described in this thesis were expressed from the Gateway® expression vector pDEST^{TM}17 which carries ampicillin (Amp) resistance. Amp was added to all growth media to a final concentration of 0.1 mg/mL. Chaperone 3 cells used for all protein expression additionally carry additional resistance to chloramphenicol (Cam) and spectinomycin (Spec). These antibiotics were added to all protein expression growth media at concentrations of 0.025 mg/mL and 0.1 mg/mL respectively. Stock solutions of Cam (1000x) were dissolved in absolute ethanol at 25 mg/mL and stock solutions of Spec (100x) were dissolved in Milli-Q at 10 mg/mL. Antibiotics were filter sterilised and stored at -80 °C, and freeze-thawed a maximum of two times.

6.2.6 Culture media

Lysogeny-broth (LB) agar was prepared by dissolving LB (Lennox L) (20 g/L) and agar (15 g/L) base in Milli-Q water and sterilised by autoclaving. The LB-agar solution was then heated in a microwave oven until boiling, and left to cool in a 60 °C oven. The LB-agar solution was then further cooled to 50 °C before antibiotic(s) were added immediately prior to pouring into round petri dishes. LB media for plasmid extraction, pre-cultures and protein expression cultures were prepared by dissolving 20 g/L LB (Lennox L) base in Milli-Q water and sterilised by autoclaving. Antibiotic(s) were added immediately prior to use.

6.3 Cloning and transformation

6.3.1 Primer design

All DNA primers used in this study were designed either by hand or using PrimerX (www.bioinformatics.org/primerx), and synthesised by Invitrogen. Lyophilized primers were re-suspended in sterilised TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) to a final concentration of 100 μM.

6.3.2 Gateway® cloning technology

All cloning and sub-cloning of genes was completed using Gateway® cloning technology (Life Technologies). The Gateway® system allows sub-cloning between expression systems without the use of traditional ligation sites and restriction enzymes. This system requires the installation of generic Gateway® 5’ and 3’ flanking sequences using sequence specific Gateway® primers. These 5’ and 3’ add-ons contain *att-B* sites which include recognition

sequences for the enzymes which catalyse their specific recombination into Gateway® donor vectors.

The gene containing generic gateway sequences is then incubated with a Gateway® donor vector (dsDNA plasmid) and an enzyme mixture known as BP clonase® II. This combination of enzymes catalyses the recombination and insertion of the gene containing generic *att-B* sites, into *att-P* recognition sites present in the donor vector sequence (BP reaction). The gene of interest, inserted into the donor vector is referred to as the entry vector. The Gateway® donor vector pDONR^{TM}221 was used for all initial cloning steps.

Once ligated into the entry vector, the target gene can be easily transferred into any Gateway® destination vector. All Gateway® destination vectors contain *att-R* recombination sequences and the enzyme mixture LR clonase® II is used to catalyse the insertion of the target gene, from entry vector, to destination vector (LR reaction). Expression of the target gene takes place from the destination vector. Destination vectors typically include additional elements such as inducible promoters and purification affinity-tag sequences, which are coupled to the target gene to facilitate selective expression and purification.

The destination vector pDEST^{TM}17 was used for expression of all proteins during this study. The pDEST^{TM}17 plasmid vector couples the gene of interest to an inducible T7 bacteriophage promoter site, and an N-terminal histidine-tag (His-tag) for purification with immobilised metal affinity chromatography (IMAC). The pDEST^{TM}17 plasmid vector carries ampicillin resistance, which is used as a selective marker for cells expressing the gene of interest.

6.3.3 Sub-cloning wild-type HsaNeuNAc 9-PS

An *E. coli* codon optimised synthetic gene encoding *Hsa*NeuNAc 9-PS was purchased from Life Technologies. This gene was purchased pre-ligated into the entry vector pDONR^{TM}221. Lyophilized pDONR^{TM}221 bearing *Hsa*NeuNAc 9-PS gene (5 µg dry weight) was re-suspended with 50 µl sterilised TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) to a final concentration of 100 ng/µl The LR reaction was then used to transfer the gene of interest into pDEST^{TM}17 destination vector for expression. For LR reaction, 1 µl of entry clone (pDONR^{TM}221), 1 µl pDEST^{TM}17 vector and 6 µl TE buffer were incubated with 2 µl LR Clonase® II mixture at room temperature for 1 hour. The reaction mixture was then incubated with 1 µl Proteinase K for 10 minutes at 37 ˚C. The reaction product (pDEST^{TM}17

vector bearing *Hsa*NeuNAc 9-PS gene) was used for transformation into competent plasmid propagation cell lines (Section 6.3.8).

6.3.4 PCR protocol

All polymerase chain reactions (PCR) were completed using a Veriti®96-well Thermal Cycler (Applied Biosystems). PCR was performed using *PfuUltra* High-Fidelity DNA Polymerase (Agilent). A 50 µL reaction mixture was made for each reaction using the manufacturer's recommended protocol (unless otherwise stated) as shown in Table 6.1. The thermal cycling parameters used are outlined in Table 6.2. PCR products were purified where necessary, using E-Gel® CloneWell™ gels (Life Technologies).

Table 6.1. PCR reaction mixtures for *PfuUltra* DNA polymerase.

PfuUltra Reaction components	Volume
Reaction buffer (commercial)	5 µL
Autoclaved milli-Q	40.6 µL
dNTP mix (25 mM each dNTP)	0.4 µL
DNA template (100 ng/µL)	1 µL
Forward primer (100 ng/µL)	1 µL
Reverse primer (100 ng/µL)	1 µL
PfuUltra DNA Polymerase (2.5U/ µL)	1 µL
Total reaction volume	50 µL

Table 6.2. Cycling parameters used for *PfuUltra* DNA polymerase.

Cycle step	Temp	Time	Cycles
Initial denaturation	95 ˚C	2 minutes	1
Denaturation, annealing and extension	95 ˚C	30 seconds	30
	Primer T_m - 5 ˚C	30 seconds	
	72 ˚C	1 minute	
Final extension	72 ˚C	10 minutes	1

6.3.5 Chimera cloning

The gene constructs encoding the $\beta_2\alpha_2$ loop-swap chimeras were generated using a 7-step overlap extension PCR methodology outlined in section 3.4. All PCR steps were completed using *PfuUltra* high-fidelity DNA polymerase according to the manufacturer's protocol (Section 6.3.4). Forward and reverse primers encoding the loop residues to be installed were designed and used to initially extend pre and post-loop fragments from the appropriate templates. The resulting linear fragments were then further extended with 'bridging primers' to introduce a 30 bp overlap between pre- and post-loop fragments. The overlapping fragments were then extended using nested gene primers containing generic gateway sequences to generate the complete linear constructs, with novel loop sequences installed. All primers used for *Nme/Hsa* $\beta_2\alpha_2$ loop-swap and double-chimera cloning are shown in table 6.3, and those used for cloning LegS and PseS loop-swap chimeras are shown in table 6.4. The *Nme/Hsa* AFPL domain-swap chimera was designed and cloned by Dr. Joseph as reported elsewhere.[86]

Table 6.3. DNA primers (5'-3') used for cloning *Nme/Hsa* chimeras. Start codons are shown in blue, stop codons in red. Overlapping sequences for generic gateway extension are underlined. Sequence corresponding to *Nme*NeuNAcS is highlighted grey, and *Hsa*NeuNAc 9-PS in green. *Att-B* sites for gateway recombination are shown in orange.

	Nme/Hsa loop-swap chimera	*Nme/Hsa* double-chimera
Template used	Wild-type *Nme*NeuNAcS pDEST^{TM}17	AFPL chimera construct pDEST^{TM}17
Forward nested gene primer	GGCAGCGGCGGCATGCAAAACAACAACGA	GGCAGCGGCGGCATGCAAAACAACAACGA
Reverse nested gene primer	GAAAGCTGGGTGTTATTCAATATCAGT	GAAAGCTGGGTGTTAGCTTTTGATTTTTTTGCC
Forward loop primer	AAACATTCATGGGGTAAAACCTATGAAATTATGGAACGTTGC	AAACATTCATGGGGTAAAACCTATGAAATTATGGAACGTTGC
Reverse loop primer	GCTGGTATACGGACGTTCCAGGGCCTCATCAGACATTTCGTC	GCTGGTATACGGACGTTCCAGGGCCTCATCAGACATTTCGTC
Forward bridging primer	CGTCCGTATACCAGCAAACATTCATGGGGT	CGTCCGTATACCAGCAAACATTCATGGGGT
Reverse bridging	ACCCCATGAATGTTTGCTGGTATACGGACG	ACCCCATGAATGTTTGCTGGTATACGGACG

primer		
Forward generic gateway primer	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAACCTGTATTTTCAGGGCAGCGGCGGC	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAACCTGTATTTTCAGGGCAGCGGCGGC
Reverse generic gateway primer	GGGGACCACTTTGTACAAGAAAGCTGGGT	GGGGACCACTTTGTACAAGAAAGCTGGGT

Table 6.4. DNA primers (5'-3') used for cloning LegS and PseS loop-swap chimeras. Start codons are shown in blue, stop codons in red. Overlapping sequences for generic gateway extension are underlined. Sequence corresponding to *Cje*NeuNAcS is highlighted grey, *Cje*LegS in yellow and *Cje*PseS purple. *Att-B* sites for gateway recombination are shown in orange.

	*Cje*LegS loop-swap chimera	*Cje*PseS loop-swap chimera
Template used	*Cje*LegS (Cj1327) pDEST^{TM}17	*Cje*PseS (Cj1317) pDEST^{TM}17
Forward nested gene primer	GGCAGCGGCGCGATGAAAAAAACTTTAATCAT	GGCAGCGGCGCGATGCAAATAGGAAATTTTAACACCG
Reverse nested gene primer	GAAAGCTGGGTGTTACTCACGGATAAGCTCATCTTCT	GAAAGCTGGGTGTCATTGGAAATCTCCTTGTTTTAAA
Forward loop primer	AATGCAAATATTAGCATTCTTCAAATGGTGCAAAAACTC	AATGCAAATATTAGCATTTATGAGCTTTATGAAAGTGCA
Reverse loop primer	ACCCGGTATAACATTTTTAGCCTTTGTGCTGATACAGTT	ACCCGGTATAACATTTTTATCGCTATTTAGTGTGAGGCT
Forward bridging primer	AATGTTATACCGGGTAATGCAAATATTAGC	AATGTTATACCGGGTAATGCAAATATTAGC
Reverse bridging primer	GCTAATATTTGCATTACCCGGTATAACATT	GCTAATATTTGCATTACCCGGTATAACATT
Forward generic gateway primer	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAACCTGTATTTTCAGGGCAGCGGCGGC	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAACCTGTATTTTCAGGGCAGCGGCGGC
Reverse generic gateway primer	GGGGACCACTTTGTACAAGAAAGCTGGGT	GGGGACCACTTTGTACAAGAAAGCTGGGT

The completed linear constructs encoding the respective chimeras were then ligated into the Gateway® entry vector pDONR^{TM}221 using the Gateway® BP reaction. 1 µl pDONR^{TM}221,

1 µl loop swap PCR product and 6 µl sterilised TE buffer were incubated with 2 µl BP clonase® II enzyme mixture at 25 ˚C for one hour. The reaction mixture was incubated with 1 µl Proteinase K for an additional 10 minutes at 37 ˚C to terminate the reaction and then stored on ice until required for transformation.

The BP reaction products (pDONR^{TM}221 vectors bearing genes of interest) were transformed into ONE SHOT® TOP10 *E. coli* plasmid propagation cells (Life Technologies) for plasmid propagation and purification (Section 6.3.8). Following purification of plasmid from pre-cultured ONE SHOT® TOP10 cells, the chimera gene constructs were verified by sequencing and transferred into expression vector pDEST^{TM}17 using the LR reaction as previously described in section 6.3.3.

6.3.6 Site-directed mutagenesis

Variants of *Hsa*NeuNAc 9-PS were generated by site-directed mutagenesis. Primers used for mutagenesis are listed in Table 6.5. All mutagenesis was completed using the pDEST^{TM}17 bearing *Hsa*NeuNAc 9-PS vector as a template, with *PfuUltra* DNA polymerase protocol (Section 6.3.4).

Table 6.5. Mutagenic primer sequences. Substituted codons are underlined.

Oligonucleotide	Primer sequence (5'- 3')
*Hsa*K290A (Fwd)	CCGTGTGAAATGGCCTGTAATGAAGCACTGGGTAAAAGCGT
*Hsa*K290A (Rev)	ACGCTTTTACCCAGTGCTTCATTACAGGCCATTTCACACGG
*Hsa*K74A (Fwd)	GGAACGTCCGTATACCAGCGCACATTCATGGGGTAAAACC
*Hsa*K74A (Rev)	GGTTTTACCCCATGAATGTGCGCTGGTATACGGACGTTCC
*Hsa*T80A (Fwd)	CATTCATGGGGTAAAGCGTATGGTGAACATAAA
*Hsa*T80A (Rev)	TTTATGTTCACCATACGCTTTACCCCATGAATG

6.3.7 Agarose gel electrophoresis

Agarose gel electrophoresis was used to ensure PCR and mutagenesis products were successfully amplified and of the correct molecular weight. 30 mL of agarose gel was prepared by combining 0.3 g agarose, 29.4 mL milli-Q water and 600 µl 50x TAE buffer (2 M Tris, 50 mM EDTA, 17.5% acetic acid). Agarose was dissolved by microwaving, then left to cool to 60˚C before adding 3 µl SYBR Safe DNA gel stain (10000x concentrate in

DMSO). The DNA gel was poured and refrigerated until set. 10 µl of each PCR product and 2 µl loading dye was loaded onto the gel and run at 85 V for 50 minutes. The running buffer used was prepared by making 5 mL TAE buffer (50x) up to 250 mL with milli-Q water.

6.3.8 Transformation into plasmid propagation cell lines and plasmid purification

All pDEST^{TM}17 or pDONR^{TM}221 vectors containing genes of interest were initially transformed into competent *E. coli* ONE SHOT® TOP10 cells for plasmid propagation. For transformation, 2 µl of vector was added to 100 µl of ONE SHOT® TOP10 cells and incubated on ice for 20 minutes. Cells were then heat-shocked at 42 ˚C for 30 seconds and returned to ice for 2 minutes. 250 µl of SOC media (2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 20 mM glucose) was added, and the transformation mixture incubated at 37 ˚C for one hour. 100 µl and 20 µl aliquots of the incubated mixture were then plated on LB agar, supplemented with 0.1 mg/mL amp. Plates were incubated overnight at 37 ˚C. Colonies from the transformation plates were selected and used to inoculate 5 mL LB cultures supplemented with ampicillin, which were then incubated overnight whilst shaking at 180 rpm, 37 ˚C in preparation for plasmid extraction.

5 mL of pre-cultured ONE SHOT® TOP10 cells bearing propagated vector were harvested by centrifugation (13000 g, 1 minute) and the cell pellet used for plasmid extraction. Purification of the propagated vector from ONE SHOT® TOP10 cells was completed using a High-Pure Plasmid Isolation kit (Roche), according to the manufacturer's instructions. Following extraction, the concentration of plasmid was measured by absorbance at 260 nm, and then vectors were sequenced to ensure the gene construct of interest was complete. Purified pDONR^{TM}221 was then used for LR reaction (Section 6.3.3), while purified pDEST^{TM}17 was used for transformation into expression cell lines.

6.3.9 Transformation into expression cell lines

Purified pDEST^{TM}17 was used for transformation into competent expression cell-lines using the same method detailed in section 6.3.8.

6.3.10 Colony PCR

After transformation of *E. coli* ONE SHOT® TOP10 cells with pDONR^{TM}221 plasmids, promising candidates for sequencing were identified with colony PCR. This procedure uses the forward gene specific DNA sequencing primer and reverse T7 terminator sequencing primer (`GCTAGTTATTGCTCAGCGG`) to confirm that the sub-cloned gene had been inserted

into the vector in the correct orientation. This was confirmed by the amplification of an appropriately sized DNA fragment from the colony PCR - subsequently visualised on agarose gel.

PCR reactions were set up using *Taq* DNA Polymerase (Invitrogen) and following the manufacturer's instructions for reaction components and cycling parameters. Template DNA was provided as a toothpick scraping of an *E. coli* TOP10 colony from an agar plate of suitably-transformed cells.

6.3.11 Glycerol stock preparation

All *E. coli* cells transformed with plasmid vector were stored as glycerol stocks for use throughout this study. 250 µl 60 % (w/w) glycerol was added to 750 µl of the pre-cultured cells and flash frozen in liquid nitrogen prior to storage at -80˚C.

6.3.12 DNA sequencing

All recombinant expression vectors were sequenced prior to transformation into expression cell lines to ensure the gene constructs of interest were complete and free of unintended mutations. All sequencing was completed at the University of Canterbury School of Biological Sciences sequencing facility using an Applied Biosystems 3130xl Genetic Analyzer. Plasmid samples sent for sequencing were approximately 50 ng/µl. Forward and reverse T7 primers at a final concentration of 3.2 µM were used for sequencing pDEST^{TM}17 vectors, while M13 primers were used for sequencing pDONR^{TM}221 vectors. Sequencing results were analysed using CLC sequence viewer 6 (Qiagen).

6.4 Protein expression and purification

6.4.1 Expression cell lines

Expression of all recombinant proteins was tested in both *E. coli* BL21 (DE3) Star cells and/or *E.coli* BL21 (DE3) pBB540/pBB542 (Chaperone 3) cells. The highest yield of soluble protein was achieved for all proteins described in this study, when expressed in chaperone 3 cells and so this cell-line was used for all up-scaled expression.

6.4.2 Protein expression for purification

Scrapings from chaperone 3 cell glycerol stocks were used to inoculate 100 mL LB pre-cultures supplemented with appropriate antibiotics (Section 6.2.5). 100 mL pre-cultures were

incubated at 37 ˚C overnight whilst shaking at 180 rpm. The pre-cultured cells were then used to inoculate 2 L of LB, and achieve a starting optical density (OD) of ~0.1 at 600 nm. 2 L cultures were grown at 37 ˚C until OD_{600} was approximately 0.4, at which time the growth temperature was reduced to 23 ˚C. When OD_{600} reached 0.5-0.6, over-expression of protein was induced by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cell culture grown overnight at 23 ˚C. Cells were harvested by centrifugation in 1 L bottles (14000 g, 15 minutes) and the cell pellet stored at -80˚C until required for purification.

6.4.3 Cell lysis

Harvested cell pellets were re-suspended in 50 mL low-salt buffer (500 mM NaCl, 50 mM K_2HPO_4, pH 8), and lysed on ice using sonication. Lysis was achieved with four rounds of sonication (35 pulse, 70-80 power, for 5 minutes) using an Omni Ruptor 4000 ultrasonic homogenizer.

For smaller volume cultures (i.e 5 mL), including those used for initial expression testing, lysis was achieved using Bugbuster® protein extraction reagent (Novagen) instead of sonication. Cell pellets were resuspended in a smaller volume of buffer (approximately 1 mL) containing Bugbuster® reagent. 0.5 µL of Benzonase nuclease® (Novagen) was added to break down DNA, and the solution shaken at 25 °C for 20-30 min.

Lysed cells were then centrifuged (21000 g, 40 minutes) to separate soluble and insoluble fractions. The soluble fraction was retained and kept on ice prior to purification.

6.4.4 Purification protocol

All proteins described in this study were purified using a common method, involving rounds of immobilised metal-affinity chromatography (IMAC) and size-exclusion chromatography. All chromatography steps were completed using a BioRad Biologic DuoFlow Chromatography system with QuadTec UV-Vis detector. Protein elution was monitored at 280 nm throughout purification. Given the range of recombinant proteins expressed purified in this study, the same buffers could not always be used. The following buffers were used for purification of the majority of wild-type and variant proteins:

Low-salt buffer (Talon equilibration): 500 mM NaCl, 50 mM K_2HPO_4 (pH 8).

High-salt buffer (Talon elution): 500 mM NaCl, 50 mM K_2HPO_4 (pH 8), 200 mM imidazole.

Size-exclusion buffer: 20 mM BTP (pH 7.5), 150 mM NaCl.

6.4.5 Immobilised metal-affinity chromatography

The soluble crude lysate was filtered and loaded onto a Tricorn 10/50 Talon immobilized metal (cobalt) affinity chromatography (IMAC) column. IMAC is used for isolation of recombinant proteins with N or C-terminal poly-histidine tags. The six histidine residues of the tag selectively chelate Co^{2+} ions immobilized within the column resin, while untagged proteins elute in the flow through. The column bound His-tagged proteins can then be eluted with high concentrations of imidazole, which displaces the histidine imidazole functionality.

The Talon column was pre-equilibrated with five column volumes of low-salt buffer, prior to sample injection. The column was then washed with ten column volumes of 100 % low-salt buffer, and the bound His-tagged protein eluted with six column volumes of 100 % high-salt buffer. Sample injection was completed at 1 mL/min while all buffer steps were completed at 3 mL/min. The tagged-protein eluted as a single peak with a distinct imidazole shoulder. Elution fractions were collected, and buffer exchanged into low-salt buffer using a GE Healthcare HiPrep 26/10 desalting column (at 3 mL/min), or manually using a 15 mL Vivaspin Turbo 10K MWCO centrifugal concentrator to reduce imidazole concentration.

6.4.6 TEV protease cleavage

The desalted fractions were then incubated with recombinant tobacco etch virus (TEV) protease for one hour at 25 ˚C then overnight at 4˚C to cleave the N-terminal His-tag. TEV protease selectively hydrolyses the polypeptide chain between glutamine and glycine of the TEV cleavage consensus sequence (Glu-Asn-Leu-Tyr-Phe-Gln/Gly) which links the His-tag and the target protein sequence. TEV protease was added at a ratio of 1:100 to the substrate protein (by mass, as determined by absorbance).

A second round of IMAC was then used to separate the cleaved target protein from free His-tags and TEV, which contains a non-cleavable His-tag. The TEV incubation mixture was re-filtered and applied to the Talon column, and run using the same protocol as for the first IMAC step (Section 6.4.5). The cleaved protein eluted in flow-through fractions while TEV protease and free His-tags were eluted with high-salt buffer. The flow-through fractions were pooled and concentrated to a volume <15 mL for size-exclusion chromatography.

6.4.7 Size-exclusion chromatography

Size-exclusion chromatography was used as a final polishing step. The concentrated protein was re-filtered and applied to a Superdex 200 26/60 size-exclusion chromatography (SEC)

column for the final step of purification. The column was pre-equilibrated with one column volume of SEC buffer (as specified) at 1 mL/min prior to sample injection. The column was then washed with an additional column volume at 1 mL/min.

6.4.8 SDS-PAGE

The purification process was monitored using sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE). Samples were taken at each stage of purification, and diluted three-fold with milli-Q water. 15 µl of the diluted sample, 1 µl of 0.5 M DTT and 5 µl of Novex Bolt™ SDS sample buffer were mixed and boiled. When excessive DNA contamination was problematic, samples were incubated with 0.5 µl Benzonase® nuclease for 5 minutes prior to boiling. 10 µl of the boiled samples were loaded onto pre-cast Novex Bolt™ 10 % Bis-Tris Plus mini-gel. 11 µl of Novex pre-stained protein standard was loaded adjacent to the sample wells. The gel was run in 1x MOPS running buffer, at 165 V for 45 minutes.

The gel was then stained for ~1 hour (1 g Coomassie Blue, 400 mL methanol, 100mL glacial acetic acid 500 mL milli-Q water), then transferred to de-stain (400 mL methanol, 100 mL glacial acetic acid and 500 mL milli-Q) for ~1 hour. De-stained gels were visualised and photographed using the BioRad Gel Doc XR system.

6.4.9 Protein concentration determination

All protein concentrations (mg/mL) were determined by absorbance at 280 nm using a NanoDrop spectrophotometer. The spectrophotometer was blanked with the appropriate buffer before making absorbance measurements in triplicate. Extinction coefficients (M^{-1} cm^{-1} at 280 nm) used for absorbance measurements are listed in table 6.6. Coefficient values were calculated based on protein amino acid sequence using ProtParam tool on the ExPASy proteomics server.

Table 6.6. Calculated protein extinction coefficients. *The difference in extinction coefficient of wild-type *Hsa*NeuNAc 9-PS and point mutant variants is negligible and thus mutant extinction coefficients are not listed here.

Protein	Extinction coefficient (M^{-1} cm^{-1} at 280 nm)
*Hsa*NeuNAc 9-PS*	31400
Nme/Hsa loop-swap chimera	28880
Nme/Hsa double-chimera	23380
*Cje*LegS loop-swap chimera	7450
*Cje*PseS loop-swap chimera	17420

6.4.10 Protein concentration and buffer exchange

Purified protein was concentrated to the desired final concentration using a 15 mL Vivaspin Turbo 10K MWCO centrifugal concentrator (4000 g, variable spin times). All concentrator units were rinsed before use with Milli-Q water. Protein solutions were buffer exchanged where necessary by repeatedly concentrating in a MWCO device and diluting with the new buffer.

6.4.11 Removal of metal from proteins

For assays requiring metal free apo-enzyme, protein was incubated with 10 mM EDTA for 30 minutes at 4˚C, prior to buffer exchange into EDTA-free chelexed buffer. Alternatively, EDTA treated protein was dialysed against EDTA-free chelexed buffer using 10 kDa MWCO Slide-A-Lyzer dialysis cassettes (Thermo Scientific). Protein was dialysed against 1000-fold volume of EDTA-free buffer for 8 hours at 4˚C. The dialysis buffer was then replaced with fresh solution before continuing with dialysis overnight, with stirring to maximise solute exchange.

6.4.12 Enzyme storage

Purified protein of the desired concentration was flash frozen with liquid nitrogen as 100 µl or 200 µl aliquots, and stored at -80˚C until required.

6.5 Protein characterisation

6.5.1 Mass spectrometry

The molecular mass of all purified proteins was confirmed by mass spectrometry. Protein was buffer exchanged into 5 mM ammonium bicarbonate, and diluted to a final concentration of 1 mg/mL for analysis. All measurements were made by electrospray ionisation using a Bruker maXis 3G (Ultra High Resolution Time of Flight) Tandem Mass Spectrometer.

6.5.2 Kinetic assay

The kinetic parameters of wild-type and variant enzymes were determined using a continuous UV-visible assay monitoring the consumption of PEP at 232 nm ($\varepsilon = 2.8 \times 10^3$ M^{-1} cm^{-1}) with a Varian Cary 100 UV Visible spectrophotometer.

For *Hsa*NeuNAc 9-PS wild-type and variants, quartz cuvettes (1 mL) containing 50 mM BTP (pH 7.5), 1 mM $MgCl_2$, 0.3 mM PEP and variable concentrations of ManNAc 6-P or Man 6-P were incubated for 10 minutes at 37 ˚C. For all chimeric proteins, quartz cuvettes (1 mL) containing 50 mM BTP (pH 7.5), 1 mM $MnCl_2$ and variable concentrations of PEP and ManNAc or ManNAc 6-P were incubated for 10 minutes at 25 ˚C. The enzymatic reactions were initiated by addition of 5 µL of purified enzyme for a final concentration of 0.025-0.05 mg/mL and the initial reaction rates determined (in triplicate) by least-squares fit. Michaelis−Menten steady-state approximations were utilized to determine kinetic parameters with nonlinear fitting in GraFit (Erithacus Software). For chimera activity tests where residual activity was expected to be low, enzyme concentration was increased up to a final maximum concentration of 1 mg/mL.

6.5.3 NMR activity assay

PRESAT NMR was used for assessment of low-level enzyme activity, with a Varian 400 MHz spectrometer. 5 mM PEP, 1 mM $MgCl_2$, 15% D_2O, 20 mM BTP (pH 7.5), 100 mM NaCl and 5 mM ManNAc or ManNAc 6-P were combined to a final volume of 1 mL and added to a 5 mm NMR tube. A blank PRESAT spectra was measured (64 scans) before adding 5 µL of enzyme at 15 mg/mL to initiate the reaction. Subsequent PRESAT spectra were measured at specific hourly time intervals following addition of enzyme. Activity was qualified by the proportional disappearance of substrate proton resonances and appearance of

product resonances. Scouting experiments were run as part of each spectra measurement in order to maximise water-suppression efficiency.

6.5.4 Circular dichroism

Circular Dichroism (CD) spectroscopy was used to ensure all recombinant proteins were correctly folded following purification, and retained wild-type like secondary structure composition. All CD measurements were performed using a JASCO J-815 Spectropolarimeter. CD spectra were recorded in triplicate from 190 to 260 nm using a 0.5 nm data pitch, 1 s response and 1 nm bandwidth at 25˚C. Protein for CD was diluted to 0.03 mg/mL with milli-Q water. A 3 mL quartz cuvette with 1 cm path length was used for each measurement. Blank spectra were recorded for milli-Q water alone for each experiment before addition of protein, and subtracted from the protein spectra obtained.

6.5.5 Differential scanning fluorimetry

Protein melt temperatures were measured in the absence and presence of ligands by differential scanning fluorimetry (DSF) using an iCycler iQ5 Multicolour Real-Time PCR Detection System (Bio-Rad). Protein samples containing 0.1 mg/mL purified protein, 50 mM BTP (pH 7.5), 10 x SYPRO orange dye and the appropriate ligand or metal to a final concentration of 1 mM (where necessary) were aliquoted as 25 µL triplicates into a 96-well microplate. A blank sample was similarly prepared for each condition, with buffer in place of the protein. The melt proceeded in 0.2 °C increments from 20 - 95 °C, with a 20 s dwell time after each temperature rise. Measurements of fluorescence were made at the end of each dwell time.

For analysis, an Excel spreadsheet with custom VBA-scripted macro was created by Dr. Timothy Allison to toggle and dynamically display melt and derivative curves for each sample. The melt temperatures were calculated as the temperature of maximum inflection of the melting curve after subtracting absorbance of the blank well. Error bars associated with DSF plots represent the standard deviation between triplicate measurements.

6.5.6 Analytical gel filtration

Native protein molecular masses were determined by analytical gel filtration on a Superdex 200 10/300 column (GE Healthcare) equilibrated with 20 mM BTP (pH 7.5) containing 150 mM NaCl. The elution volume of blue dextran (2000 kDa) was designated as the column void volume, and the column was then equilibrated with 1 mg/mL of known molecular

weight standards (Sigma Aldrich) including thyroglobulin (669 kDa), beta amylase (200 kDa), alcohol dehydrogenase (150 kDa), ovalbumin (43 kDa), conalbumin (75 kDa) and carbonic anhydrase (29 kDa). Protein samples of 3 mg/mL were then run at 0.2 mL/min and the native molecular mass calculated as function of elution volume relative to elution volume of the molecular weight standards.

6.5.7 Isothermal titration calorimetry

The thermodynamic parameters of ligand binding for wild-type and variants of *Hsa*NeuNAc 9-PS were measured by isothermal titration calorimetry (ITC) using a VP-ITC unit operating at 298 K (MicroCal; GE Healthcare). Prior to use, protein was treated with 1 mM EDTA, and then buffer exchanged into chelexed 20 mM BTP (pH 7.5) with 150 mM NaCl. Stock solutions of ligand were then made using the same buffer, and pH adjusted to match the protein solution. All solutions were filtered and degassed under vacuum, and protein concentration was measured by UV absorption. For PEP binding experiments, 2 mM PEP was titrated into 90 μM protein with a background of 1 mM $MgCl_2$. For ManNAc 6-P binding experiments, 5 mM ManNAc 6-P was titrated into 90 μM protein with a background of 1 mM $MgCl_2$. Experiments were performed with a starting sample volume of 1.4 mL in the cell. Each titration proceeded via 57 injections of ligand: one 2 μL injection followed by 56 injections of 5 μL.

ITC titrations of wild-type *Nme*NeuNAcS and all chimeric proteins were completed using an ITC200 unit operating at 298 K (MicroCal; GE Healthcare) using the same sample preparation as described above however samples were not degassed in this case. For PEP binding experiments, 1 mM PEP was titrated into 150 μM protein with a background of 1 mM $MnCl_2$. Experiments were performed with a starting sample volume of 200 μL in the cell. Each titration proceeded via 20 injections of ligand: one 0.5 μL injection followed by 19 injections of 2 μL.

Heats of dilution experiments (ligand titrated into buffer + 1 mM metal) were measured independently and subtracted from the integrated data before curve-fitting in Origin 7.0 with the standard one-site model supplied by MicroCal.

6.5.8 Small-angle X-ray scattering data collection

Small angle X-ray Scattering (SAXS) measurements were performed at the Australian Synchrotron SAXS/WAXS beamline equipped with a Pilatus detector (1 M, 170 mm × 170 mm, effective pixel size of 172 μm × 172 μm). The wavelength of the X-rays was 1.033 Å. The sample–detector distance was 1575 mm, which provided a q range of 0.01–0.555$Å^{-1}$ [where q is the magnitude of the scattering vector, which is related to the scattering angle (2θ) and the wavelength (λ) as follows: $q = (4\pi/\lambda)\sin\theta$]. Data were collected from a 1.5 mm glass capillary at 25 ˚C at 2 second intervals.

Scattering data was collected for all proteins (~10 mg/ml) following elution from a size-exclusion chromatography column (Superdex 200 5/150), pre-equilibrated with 20 mM BTP (pH 7.5) with 150 mM NaCl, in both the presence and absence of 1 mM PEP and 1mM $MnCl_2$. Two-dimensional intensity plots from the peak of the size-exclusion chromatography run were radially averaged, normalized to sample transmission, and background subtracted.

6.5.9 Small-angle X-ray scattering data processing

Scattered intensity (I) was plotted *versus* q. All samples were devoid of an increase in intensity at low q (indicative of aggregation) unless otherwise mentioned. All Guinier plots were linear for $qR_g < 1.3$ unless otherwise mentioned. One-dimensional profiles were background subtracted, and Guinier fits were made using PRIMUS.[105] Fourier transform was performed using PRIMUS to yield the function P(r), which gives both the relative probabilities of distances between scattering centres and the maximum dimension of the scattering particle D_{max}. Theoretical scattering curve of *Hsa*NeuNAc 9-PS homology model was generated from atomic coordinates and compared with experimental scattering curves using CRYSOL.[78] Similarly, theoretical scattering of wild-type *Nme*NeuNAcS was generated based on crystallographic coordinates (PDB: 1XUZ) and compared with the scattering profiles obtained for the chimeric proteins using CRYSOL.

6.5.10 Crystallisation trials

Crystal screens were either completed by hanging-drop vapour diffusion using 24-well VDX plates (Hampton Research) or using a Mosquito Crystal® (TTP Labtech) crystallisation robot on a 96-well sitting drop iQ plate (TTP Labtech). For 24-well plate screens, a final drop volume of 2 μL was used, consisting of 1 μL protein solution (in SEC buffer) and 1 μL reservoir solution (total reservoir volume of 500 μL). For 96-well sitting drop plates, a final

drop volume of 400 nL (equal parts protein and condition) and reservoir volume of 40 µL was used. Crystallisation screening trays were incubated at 20 ˚C or 4 ˚C indefinitely.

Protein for crystallisation was taken from the middle two fractions of the SEC elution peak following purification, and used immediately where possible. Protein crystallisation trials were attempted in the presence and absence of $MgCl_2$/$MnCl_2$ and PEP, which was added by buffer exchanging protein into SEC buffer with additional ligands present at 1 mM.

Proteins were screened at varied concentrations between 1 and 10 mg/mL (final concentration in drop) using commercially available crystallisation screening conditions. Protein crystallisation conditions used include: Molecular Dimensions Morpheus, JCSG-plus, PACT premier, Clear Strategy I, Clear Strategy II and ProPlex HT-96 screens.

Chapter 7. Appendices

7.1 Appendix A. Sequence alignments

```
HsaNeuNAc-9-PS   MPLELEL-CPGRWVGGQHPCFIIAEIGQNHQGDLDVAKRMIRMAKECGADCAKFQKSELE   59
NmeNeuNAcS       MQNNNEFKIGNRSVGYNHEPLIICEIGINHEGSLKTAFEMVDAAYNAGAEVVKHQTHIVE   60
                 *  : *:    * ** :*  :**.*** **:*.*..* .*:  * :.**: .*.*.  :*

HsaNeuNAc-9-PS   FKFNRKALERPYTSKHSWGKTYGEHKRHLEFSHDQYRELQRYAEEVGIFFTASGMDEMAV  119
NmeNeuNAcS       DEMSDEAKQVI--PGNADVSIY-EIMERCALNEEDEIKLKEYVESKGMIFISTPFSRAAA  117
                  ::. :* :      .:  . * *  .:  :..::  :*:.*.*. *::* :: :.. *.

HsaNeuNAc-9-PS   EFLHELNVPFFKVGSGDTNNFPYLEKTAKKGRPMVISSGMQSMDTMKQVYQIVKPLNPNF  179
NmeNeuNAcS       LRLQRMDIPAYKIGSGECNNYPLIKLVASFGKPIILSTGMNSIESIKKSVEIIREAGVPY  177
                   *:.:::* :*:***: **:* :: .*. *:*:::*:**:*::::*:  :*::     :

HsaNeuNAc-9-PS   CFLQCTSAYPLQPEDVNLRVISEYQKLFPDIPIGYSGHETGIAISVAAVALGAKVLERHI  239
NmeNeuNAcS       ALLHCTNIYPTPYEDVRLGGMNDLSEAFPDAIIGLSDHTLDNYACLGAVALGGSILERHF  237
                 .:*:**. **   ***.*  :.: .: ***  ** * *      .:.*****..:****:

HsaNeuNAc-9-PS   TLDKTWKGSDHSASLEPGELAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKV  299
NmeNeuNAcS       TDRMDRPGPDIVCSMNPDTFKELKQGAHALKLARGGKKDTIIAGEKPTKDFAFASVVADK  297
                 *      * *  .*::*  : ** :..: :: * *. .. ::  *   ::    ****.

HsaNeuNAc-9-PS   KIPEGTILTMDMLTVKVGEPKGYPPEDIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS  359
NmeNeuNAcS       DIKKGELLSGDNLWVKRPGNGDFSVNEYETLFGKVAACNIRKGAQIKKTDIE--------  349
                 .* :* :*: * * **      :  ::  .*.** .  .:.:   * :  ::
```

```
1: HsaNeuNAc-9-PS  100.00  28.7   %
2: NmeNeuNAcS       28.7    100.00 %
```

Figure 7.1. Pairwise alignment of *Hsa*NeuNAc 9-PS and *Nme*NeuNAcS. Catalytic barrel residues are blue, $\beta_2\alpha_2$ loop residues red, linker region residues yellow and AFPL domain residues green. Domain annotation is based on the structure of *Nme*NeuNAcS (PDB 1XUZ).

```
CjePseS          --MQIGNF-NTDKKVFIIAELSANHAGSLEMALKSIKAAKKAGADAIKIQTYTPDSLTLN   57
CjeNeuNAcS       MQIKIDKLTISQKNPLIIPEIGINHNGSLEIAKLMVDAAKRAGAKIIKHQTHIVEDEMSQ   60
CjeLegS          -----------MKKTLIIAEAGVNHNGDLNLAKKLIEIAADSGADFVKFQSFKAKNCIST   49
                            *: :** * . ** *.*::*   :. *  :**. :* *:.  ..

CjePseS          SDK-EDFIIKGGLWDKRKLYELYESAKTPYEWHSQIFETAQNEGILCFSSPFAKEDVEFL  116
CjeNeuNAcS       EAKNVI----PG-NANISIYEIMEQCALNYKDELALKEYVEKQGLVYLSTPFSRAAANRL  115
CjeLegS          KAKKAPYQLKTT-ANDESQLQMVQKLELDLKAHKELILHAKKCNIAFLSTPFDLESVDLL  108
                 . *            . .  :: :.     : .  :   .::  :  :*:**    .: *

CjePseS          KRFDPIAYKIASFEANDENFVRLIAKEKKPTIVSTGIATEEELFKICEIFKEE--KNPDL  174
CjeNeuNAcS       EDMGVSAYKIGSGECNNYPLIKHIAQFKKPMIISTGMNSIESIKPTVKILRDY---EIPF  172
CjeLegS          NELGLKIFKIPSGEITNLPYLKKIAKLNKKIILSTGMANLGEIEEALNVLCKNGAKRQNI  168
                 : :    :** * * .:   :: **: :*  *:***: .  .:    ::: .    .  :

CjePseS          VFLKCTSTYPTAIEDMNLKGIVSLKEKFNVEVGLSDHSFGFLAPVMAVALGARVIEKHFM  234
CjeNeuNAcS       VLLHTTNLYPTPSHLVRLQAMLELYKEFNCLYGLSDHTTNNLACIGAIALGASVLERHFT  232
CjeLegS          TLLHCTTEYPAPFNEVNLKAMQSLKDAFKLDVGYSDHTRGIHISLAAVALGACVIEKHFT  228
                 .:*: *. **:  . :.*:.: .* . *:   * ***:      : *:**** *:*:**

CjePseS          LDKSIESEDSKFSLDFDEFKAMVDAVRQAESALGDGKLDLDEKVLK-NRVFARSLYASKD  293
CjeNeuNAcS       DTMDRKGPDIVCSMDESTLKDLINQTQEMVLLRGDNNKNPLKEEQVTIDFAFASVVSIKD  292
CjeLegS          LDKNMSGPDHKASLEPQELKMLCTQIRQIQKAMGDGIKKASKSEQKNINIVRKSLVAKKD  288
                   . .. *   *:: . :* :    ::     **   .  :.      .   *: : **

CjePseS          IKKGEMFSEENVKSVRPSF-GLHPKFYQELLGKKASKDIKFGDALKQGDFQ           343
CjeNeuNAcS       IKKGEILSMDNIWVKRPSKGGISAKDFEAILGKRAKKDIKNNIQLTWDDFE           343
CjeLegS          IKKGEIFSEGNLTTKRPAN-GISAMRYEEFLGKIATKNYKEDELIRE----           334
                 *****::*  *:   **:  *:    :: :*** *.*: *    :
```

PERCENTAGE IDENTITY MATRIX:

1: *Cje*PseS	100.00	32.6	35.2	%
2: *Cje*NeuNAcS	32.6	100.00	36.7	%
3: *Cje*LegS	35.2	36.7	100.00	%

Figure 7.2. Multiple alignment of *C. jejuni* sialic acid synthases. Catalytic barrel residues are blue, $\beta_2\alpha_2$ loop residues red, linker region residues yellow and AFPL domain residues green. Domain annotation is based on the structure of *Nme*NeuNAcS (PDB 1XUZ).

7.2 Appendix B. Mass spectrometry derived molecular weights

Protein	Experimental mass (Da)	Theoretical mass from sequence (Da)	Error (%)
Wild-type HsaNeuNAc 9-PS	40494.6	40495.7	0.003
*Hsa*K290A	40436.8	40437.5	0.002
*Hsa*K74A	40437.5	40437.5	0.00
*Hsa*T80A	40465.1	40465.7	0.002
Nme/Hsa loop-swap chimera	39067.3	39069.4	0.005
Nme/Hsa double-chimera	40020.8	40022.9	0.005
LegS loop-swap chimera	36780.2	36781.0	0.002
PseS loop-swap chimera	38211.5	38212.6	0.003

References

1. Tanner, M. E., The enzymes of sialic acid biosynthesis. *Bioorg. Chem.* **2005,** *33* (3), 216-228.

2. Traving, C.; Schauer, R., Structure, function and metabolism of sialic acids. *Cellular and Molecular Life Sciences* **1998,** *54* (12), 1330-1349.

3. Schauer, R., Sialic acids: fascinating sugars in higher animals and man. *Zoology (Jena)* **2004,** *107* (1), 49-64.

4. Hao, J. J.; Balagurumoorthy, P.; Suryakala, S.; Sundaramoorthy, M., Cloning, expression, and characterization of sialic acid synthases. *Biochem. Biophys. Res. Commun.* **2005,** *338* (3), 1507-1514.

5. Kim, K.; Lawrence, S. M.; Park, J.; Pitts, L.; Vann, W. F.; Betenbaugh, M. J.; Palter, K. B., Expression of a functional *Drosophila melanogaster N*-acetylneuraminic acid (Neu5Ac) phosphate synthase gene: evidence for endogenous sialic acid biosynthetic ability in insects. *Glycobiology* **2002,** *12* (2), 73-83.

6. Hao, J.; Vann, W. F.; Hinderlich, S.; Sundaramoorthy, M., Elimination of 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid 9-phosphate synthase activity from human *N*-acetylneuraminic acid 9-phosphate synthase by a single mutation. *Biochem. J.* **2006,** *397*, 195-201.

7. Severi, E.; Hood, D. W.; Thomas, G. H., Sialic acid utilization by bacterial pathogens. *Microbiology-Sgm* **2007,** *153*, 2817-2822.

8. Goon, S.; Kelly, J. F.; Logan, S. M.; Ewing, C. P.; Guerry, P., Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the Cj1293 gene. *Mol. Microbiol.* **2003,** *50* (2), 659-671.

9. Schoenhofen, I. C.; McNally, D. J.; Brisson, J.-R.; Logan, S. M., Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: synthesis from UDP-N-acetylglucosamine by a single enzymatic reaction. *Glycobiology* **2006,** *16* (9), 8C-14C.

10. Schoenhofen, I. C.; Vinogradov, E.; Whitfield, D. M.; Brisson, J.-R.; Logan, S. M., The CMP-legionaminic acid pathway in *Campylobacter:* Biosynthesis involving novel GDP-linked precursors. *Glycobiology* **2009,** *19* (7), 715-725.

11. Linton, D.; Karlyshev, A. V.; Hitchen, P. G.; Morris, H. R.; Dell, A.; Gregson, N. A.; Wren, B. W., Multiple *N*-acetyl neuraminic acid synthetase (neuB) genes in *Campylobacter jejuni*: identification and characterization of the gene involved in sialylation of lipo-oligosaccharide. *Mol. Microbiol.* **2000,** *35* (5), 1120-1134.

12. Munday, J.; Floyd, H.; Crocker, P. R., Sialic acid binding receptors (siglecs) expressed by macrophages. *J. Leukoc. Biol.* **1999,** *66* (5), 705-711.

13. Fukuda, M., Possible roles of tumor-associated carbohydrate antigens. *Cancer Res.* **1996,** *56* (10), 2237-2244.

14. Seidenfaden, R.; Krauter, A.; Schertzinger, F.; Gerardy-Schahn, R.; Hildebrandt, H., Polysialic acid directs tumor cell growth by controlling heterophilic neural cell adhesion molecule interactions. *Mol. Cell. Biol.* **2003,** *23* (16), 5908-5918.

15. Wang, B.; Miller, J. B.; McNeil, Y.; McVeagh, P., Sialic Acid Concentration of Brain Gangliosides: Variation Among Eight Mammalian Species. *Comp. Biochem. Physiol., A: Comp. Physiol.* **1998,** *119* (1), 435-439.

16. Ferrari, G.; Batistatou, A.; Greene, L. A., Gangliosides rescue neuronal cells from death after trophic factor deprivation. *J. Neurosci.* **1993,** *13* (5), 1879-87.

17. Modi, P.; Sadasivudu, B.; Lakshminarayana, U.; Murthy, C. R., Functional relationship between ammonia and gangliosides in brain. *Neurochem. Res.* **1994,** *19* (3), 353-8.

18. Wang, B., Sialic acid is an essential nutrient for brain development and cognition. *Annu. Rev. Nutr.* **2009,** *29*, 177-222.

19. Schwarzkopf, M.; Knobeloch, K. P.; Rohde, E.; Hinderlich, S.; Wiechens, N.; Lucka, L.; Horak, I.; Reutter, W.; Horstkorte, R., Sialylation is essential for early development in mice. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (8), 5267-70.

20. Vimr, E. R.; Kalivoda, K. A.; Deszo, E. L.; Steenbergen, S. M., Diversity of Microbial Sialic Acid Metabolism. *Microbiol. Mol. Biol. Rev.* **2004,** *68* (1), 132-153.

21. Astronomo, R. D.; Burton, D. R., Carbohydrate vaccines: developing sweet solutions to sticky situations? *Nature Reviews Drug Discovery* **2010,** *9* (4), 308-324.

22. Weis, W.; Brown, J. H.; Cusack, S.; Paulson, J. C.; Skehel, J. J.; Wiley, D. C., Structure of the influenza-virus hemagglutinin complexed with its receptor, sialic-acid. *Nature* **1988,** *333* (6172), 426-431.

23. Von Itzstein, M., The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discov.* **2007,** *6* (12), 967-974.

24. Gong, J. Z.; Xu, W. F.; Zhang, J., Structure and functions of influenza virus neuraminidase. *Curr. Med. Chem.* **2007,** *14* (1), 113-122.

25. Varghese, J. N., Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Dev. Res.* **1999,** *46* (3-4), 176-196.

26. Comb, D. G.; Roseman, S., Composition and enzymatic synthesis of *N*-acetylneuraminic acid (sialic acid). *J. Am. Chem. Soc.* **1958,** *80* (2), 497-499.

27. Roseman, S.; Rood, R.; Watson, D.; Jourdian, G. W., Enzymatic synthesis of sialic acid 9-phosphates. *Proc. Natl. Acad. Sci. U. S. A.* **1961,** *47* (7), 958-961.

28. Blacklow, R. S.; Warren, L., Biosynthesis of sialic acids by *Neisseria meningitidis*. *J. Biol. Chem.* **1962,** *237* (11), 3520-3526.

29. Warren, L.; Felsenfeld, H., Biosynthesis of sialic acids. *J. Biol. Chem.* **1962,** *237* (5), 1421-1431.

30. Chou, W. K.; Hinderlich, S.; Reutter, W.; Tanner, M. E., Sialic acid biosynthesis: Stereochemistry and mechanism of the reaction catalyzed by the mammalian UDP-N-acetylglucosamine 2-epimerase. *J. Am. Chem. Soc.* **2003,** *125* (9), 2455-2461.

31. Bravo, I. G.; García-Vallvé, S.; Romeu, A.; Reglero, A., Prokaryotic origin of cytidylyltransferases and α-ketoacid synthases. *Trends Microbiol.* **2004,** *12* (3), 120-128.

32. Gunawan, J.; Simard, D.; Gilbert, M.; Lovering, A. L.; Wakarchuk, W. W.; Tanner, M. E.; Strynadka, N. C. J., Structural and mechanistic analysis of sialic acid synthase NeuB from *Neisseria meningitidis* in complex with Mn^{2+} phosphoenolpyruvate, and *N*-acetylmannosaminitol. *J. Biol. Chem.* **2005,** *280* (5), 3555-3563.

33. Chen, H.; Blume, A.; Zimmermann-Kordmann, M.; Reutter, W.; Hinderlich, S., Purification and characterization of *N*-acetylneuraminic acid-9-phosphate synthase from rat liver. *Glycobiology* **2002,** *12* (2), 65-71.

34. Mizanur, R. M.; Pohl, N. L., Bacterial CMP-sialic acid synthetases: production, properties, and applications. *Appl. Microbiol. Biotechnol.* **2008,** *80* (5), 757-765.

35. Inoue, S.; Kitajima, K., KDN (deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family. *Glycoconj. J.* **2006,** *23* (5-6), 277-290.

36. Nadano, D.; Iwasaki, M.; Endo, S.; Kitajima, K.; Inoue, S.; Inoue, Y., A naturally-occurring deaminated neuraminic acid, 3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN) - its unique occurrence at the nonreducing ends of oligosialyl chains in polysialoglycoprotein of rainbow-trout eggs. *J. Biol. Chem.* **1986,** *261* (25), 1550-1557.

37. Lawrence, S. M.; Huddleston, K. A.; Pitts, L. R.; Nguyen, N.; Lee, Y. C.; Vann, W. F.; Coleman, T. A.; Betenbaugh, M. J., Cloning and expression of the human *N*-acetylneuraminic acid phosphate synthase gene with 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid biosynthetic ability. *J. Biol. Chem.* **2000,** *275* (23), 17869-17877.

38. Angata, T.; Kitazume, S.; Terada, T.; Kitajima, K.; Inoue, S.; Troy, F. A.; Inoue, Y., Identification, characterization, and developmental expression of a novel alpha-2- 8-KDN-transferase which terminates elongation of alpha-2- 8-linked oligo-polysialic acid chain synthesis in trout egg polysialoglycoproteins. *Glycoconj. J.* **1994,** *11* (5), 493-499.

39. Knirel, Y. A.; Kocharova, N. A.; Shashkov, A. S.; Kochetkov, N. K.; Mamontova, V. A.; Soloveva, T. F., Structure of the capsular polysaccharide of *Klebsiella ozaenae* serotype-k4 containing 3-Deoxy-D-glycero-D-galacto-nonulosonic acid. *Carbohydr. Res.* **1989,** *188*, 145-155.

40. Inoue, S.; Lin, S. L.; Chang, T. N.; Wu, S. H.; Yao, C. W.; Chu, T. Y.; Troy, F. A.; Inoue, Y., Identification of free deaminated sialic acid (2-keto-3-deoxy-D-glycero-D-galacto-nononic acid) in human red blood cells and its elevated expression in fetal cord red blood cells and ovarian cancer cells. *J. Biol. Chem.* **1998,** *273* (42), 27199-27204.

41. Angata, T.; Nakata, D.; Matsuda, T.; Kitajima, K.; Troy, F. A., Biosynthesis of KDN (2-keto-3-deoxy-D-glycero-D-galacto-nononic acid) - Identification and characterization of a KDN-9-phosphate synthetase activity from trout testis. *J. Biol. Chem.* **1999,** *274* (33), 22949-22956.

42. Glaze, P. A.; Watson, D. C.; Young, N. M.; Tanner, M. E., Biosynthesis of CMP-N,N'-Diacetyllegionaminic acid from UDP-N,N'-Diacetylbacillosamine in *Legionella pneumophila*. *Biochemistry* **2008,** *47* (10), 3272-3282.

43. Cianciotto, N. P., Pathogenicity of Legionella pneumophila. *Int. J. Med.l Microbiol.* **2001,** *291* (5), 331-343.

44. McNally, D. J.; Aubry, A. J.; Hui, J. P. M.; Khieu, N. H.; Whitfield, D.; Ewing, C. P.; Guerry, P.; Brisson, J. R.; Logan, S. M.; Soo, E. C., Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *J. Biol. Chem.* **2007,** *282* (19), 14463-14475.

45. Berg, H. C., The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **2003,** *72*, 19-54.

46. Thibault, P.; Logan, S. M.; Kelly, J. F.; Brisson, J. R.; Ewing, C. P.; Trust, T. J.; Guerry, P., Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **2001,** *276* (37), 34862-70.

47. Duan, Q. D.; Zhou, M. X.; Zhu, L. Q.; Zhu, G. Q., Flagella and bacterial pathogenicity. *J. Basic Microbiol.* **2013,** *53* (1), 1-1.

48. Chou, W. K.; Dick, S.; Wakarchuk, W. W.; Tanner, M. E., Identification and characterization of NeuB3 from *Campylobacter jejuni* as a pseudaminic acid synthase. *J. Biol. Chem.* **2005,** *280* (43), 35922-35928.

49. Ewing, C. P.; Andreishcheva, E.; Guerry, P., Functional Characterization of Flagellin Glycosylation in *Campylobacter jejuni* 81-176. *J. Bacteriol.* **2009,** *191* (22), 7086-7093.

50. Schirm, M.; Soo, E. C.; Aubry, A. J.; Austin, J.; Thibault, P.; Logan, S. M., Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **2003,** *48* (6), 1579-92.

51. Logan, S. M.; Kelly, J. F.; Thibault, P.; Ewing, C. P.; Guerry, P., Structural heterogeneity of carbohydrate modifications affects serospecificity of *Campylobacter* flagellins. *Mol. Microbiol.* **2002,** *46* (2), 587-597.

52. Buschiazzo, A.; Alzari, P. M., Structural insights into sialic acid enzymology. *Curr. Opin. Chem. Biol.* **2008,** *12* (5), 565-572.

53. Vann, W. F.; Tavarez, J. J.; Crowley, J.; Vimr, E.; Silver, R. P., Purification and characterization of the Escherichia coli Kl neuB gene product N-acetylneuraminic acid synthetase. *Glycobiology* **1997,** *7* (5), 697-701.

54. Sundaram, A. K.; Pitts, L.; Muhammad, K.; Wu, J.; Betenbaugh, M.; Woodard, R. W.; Vann, W. F., Characterization of N-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. *Biochem. J.* **2004,** *383*, 83-89.

55. Suryanti, V.; Nelson, A.; Berry, A., Cloning, over-expression, purification, and characterisation of *N*-acetylneuraminate synthase from *Streptococcus agalactiae*. *Protein Expr. Purif.* **2003,** *27* (2), 346-356.

56. Liu, F.; Lee, H. J.; Strynadka, N. C. J.; Tanner, M. E., Inhibition of *Neisseria meningitidis* Sialic Acid Synthase by a Tetrahedral Intermediate Analogue. *Biochemistry* **2009,** *48* (39), 9194-9201.

57. Vimr, E. R.; Aaronson, W.; Silver, R. P., Genetic-analysis of chromosomal mutations in the polysialic acid gene-cluster of *Escherichia-coli* k1. *J. Bacteriol.* **1989,** *171* (2), 1106-1117.

58. Annunziato, P. W.; Wright, L. F.; Vann, W. F.; Silver, R. P., Nucleotide-sequence and genetic-analysis of the NeuD and NeuB genes in region-2 of the polysialic acid gene-cluster of *Escherichia coli*-k1. *J. Bacteriol.* **1995,** *177* (2), 312-319.

59. Ganguli, S.; Zapata, G.; Wallis, T.; Reid, C.; Boulnois, G.; Vann, W. F.; Roberts, I. S., Molecular-cloning and analysis of genes for sialic-acid synthesis in *Neisseria meningitidis* group-B and purification of the meningococcal Cmp-NeuNAc synthetase enzyme. *J. Bacteriol.* **1994,** *176* (15), 4583-4589.

60. Hwang, T. S.; Hung, C. H.; Teo, C. F.; Chen, G. T.; Chang, L. S.; Chen, S. F.; Chen, Y. J.; Lin, C. H., Structural characterization of Escherichia coli sialic acid synthase. *Biochem. Biophys. Res. Commun.* **2002,** *295* (1), 167-173.

61. Huang, H. H.; Liao, H. K.; Chen, Y. J.; Hwang, T. S.; Lin, Y. H.; Lin, C. H., Structural characterization of sialic acid synthase by electrospray mass spectrometry - A tetrameric enzyme composed of dimeric dimers. *J. Am. Soc. Mass Spectrom.* **2005,** *16* (3), 324-332.

62. Baardsnes, J.; Davies, P. L., Sialic acid synthase: the origin of fish type III antifreeze protein? *Trends Biochem. Sci.* **2001,** *26* (8), 468-469.

63. Salvay, A. G.; Gabel, F.; Pucci, B.; Santos, J.; Howard, E. I.; Ebel, C., Structure and interactions of fish type III antifreeze protein in solution. *Biophys. J.* **2010,** *99* (2), 609-18.

64. Joseph, D. D. A.; Jiao, W.; Parker, E. J., Arg314 Is Essential for Catalysis by *N*-Acetyl Neuraminic Acid Synthase from *Neisseria meningitidis*. *Biochemistry* **2013,** *52* (15), 2609-2619.

65. Silverman, J. A.; Balakrishnan, R.; Harbury, P. B., Reverse engineering the (beta/alpha)(8) barrel fold. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98* (6), 3092-3097.

66. Shumilin, I. A.; Bauerle, R.; Kretsinger, R. H., The high-resolution structure of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase reveals a twist in the plane of bound phosphoenolpyruvate. *Biochemistry* **2003,** *42* (13), 3766-3776.

67. Reichau, S.; Jiao, W.; Walker, S. R.; Hutton, R. D.; Baker, E. N.; Parker, E. J., Potent Inhibitors of a Shikimate Pathway Enzyme from *Mycobacterium tuberculosis* - Combining mechanism- and modeling-based design. *J. Biol. Chem.* **2011,** *286* (18), 16197-207.

68. Watson, D. R.; Jourdian, G. W.; Roseman, S., The sialic acids. 8. Sialic acid 9-phosphate synthetase. *J. Biol. Chem.* **1966,** *241* (23), 5627-36.

69. Reaves, M. L.; Lopez, L. C.; Daskalova, S. M., Replacement of the antifreeze-like domain of human *N*-acetylneuraminic acid phosphate synthase with the mouse antifreeze-like domain impacts both *N*-acetylneuraminic acid 9-phosphate synthase and 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid 9-phosphate synthase activities. *Bmb Reports* **2008,** *41* (1), 72-78.

70. Hamada, T.; Ito, Y.; Abe, T.; Hayashi, F.; Guntert, P.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yoshida, M.; Tanaka, A.; Sugano, S.; Yokoyama, S.; Hirota, H., Solution structure of the antifreeze-like domain of human sialic acid synthase. *Protein Sci.* **2006,** *15* (5), 1010-1016.

71. Spellberg, B.; Guidos, R.; Gilbert, D.; Bradley, J.; Boucher, H. W.; Scheld, W. M.; Bartlett, J. G.; Edwards, J.; America, IDSA., The Epidemic of Antibiotic-Resistant Infections: A Call to Action for the Medical Community from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **2008,** *46* (2), 155-164.

72. Risso, V. A.; Gavira, J. A.; Mejia-Carmona, D. F.; Gaucher, E. A.; Sanchez-Ruiz, J. M., Hyperstability and Substrate Promiscuity in Laboratory Resurrections of Precambrian β-Lactamases. *J. Am. Chem. Soc.* **2013,** *135* (8), 2899-2902.

73. Davies, J.; Davies, D., Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* **2010,** *74* (3), 417-433.

74. Brouwer, M. C.; Tunkel, A. R.; van de Beek, D., Epidemiology, diagnosis, and antimicrobial treatment of acute bacterial meningitis. *Clin. Microbiol. Rev.* **2010,** *23* (3), 467-92.

75. Gupta, N.; Limbago, B. M.; Patel, J. B.; Kallen, A. J., Carbapenem-Resistant Enterobacteriaceae: Epidemiology and Prevention. *Clin. Infect. Dis.* **2011,** *53* (1), 60-67.

76. Gorelik, E.; Galili, U.; Raz, A., On the Role of Cell Surface Carbohydrates and their Binding Proteins (lectins) in Tumor Metastasis. *Cancer Metastasis Rev.* **2001,** *20* (3-4), 245-277.

77. Joseph, D. D. A.; Jiao, W.; Kessans, S. A.; Parker, E. J., Substrate-mediated control of the conformation of an ancillary domain delivers a competent catalytic site for *N*-acetylneuraminic acid synthase. *Proteins Struct. Funct. Bioinformat.* **2014,** *DOI: 10.1002*, prot.24558.

78. Svergun, D.; Barberato, C.; Koch, M. H. J., CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **1995,** *28*, 768-773.

79. Fischer, H.; de Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F., Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **2010,** *43* (1), 101-109.

80. Wagner, T.; Shumilin, I. A.; Bauerle, R.; Kretsinger, R. H., Structure of 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli:* comparison of the Mn^2+2-phosphoglycolate and the Pb^2+2-Phosphoenolpyruvate complexes and implications for catalysis. *J. Mol. Biol.* **2000,** *301* (2), 389-399.

81. Schofield, L. R.; Anderson, B. F.; Patchett, M. L.; Norris, G. E.; Jameson, G. B.; Parker, E. J., Substrate Ambiguity and Crystal Structure of *Pyrococcus furiosus* 3-Deoxy-d-arabino-heptulosonate-7-phosphate Synthase: An Ancestral 3-Deoxyald-2-ulosonate-phosphate Synthase? *Biochemistry* **2005,** *44* (36), 11950-11962.

82. KEGG. (2013) Kyoto encyclopaedia of genes and genomes.

83. Goujon, M.; McWilliam, H.; Li, W.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R., A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **2010,** *38* (Web Server issue), W695-9.

84. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **2011,** *7*, 539.

85. Walker, J. E.; Saraste, M.; Runswick, M. J.; Gay, N. J., Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1982,** *1* (8), 945-51.

86. Joseph, D. D. A. Mechanistic and Evolutionary Analyses of the Sialic Acid Synthase Family. University of Canterbury, 2014.

87. Brigham, C.; Caughlan, R.; Gallegos, R.; Dallas, M. B.; Godoy, V. G.; Malamy, M. H., Sialic Acid (*N*-Acetyl Neuraminic Acid) Utilization by *Bacteroides fragilis* Requires a Novel *N*-Acetyl Mannosamine Epimerase. *J. Bacteriol.* **2009,** *191* (11), 3629-3638.

88. Ringenberg, M. A.; Steenbergen, S. M.; Vimr, E. R., The first committed step in the biosynthesis of sialic acid by *Escherichia coli* K1 does not involve a phosphorylated *N*-acetylmannosamine intermediate. *Mol. Microbiol.* **2003,** *50* (3), 961-975.

89. Svergun, D. I.; Koch, M. H. J., Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **2003,** *66* (10), 1735-1782.

90. Lewis, A. L.; Desa, N.; Hansen, E. E.; Knirel, Y. A.; Gordon, J. I.; Gagneux, P.; Nizet, V.; Varki, A., Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (32), 13552-7.

91. Vimr, E. R., Microbial sialidases: does bigger always mean better? *Trends Microbiol.* **1994,** *2* (8), 271-277.

92. Schauer, R., Sialic acids: fascinating sugars in higher animals and man. *Zoology* **2004,** *107* (1), 49-64.

93. Schrödinger, L. *The PyMOL Molecular Graphics System, Version 1.5.0.4*

94. Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J., Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009,** *25* (9), 1189-1191.

95. Rambaut, A. http://tree.bio.ed.ac.uk/software/figtree/.

96. Jacobson, M. P.; Friesner, R. A.; Xiang, Z. X.; Honig, B., On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **2002,** *320* (3), 597-608.

97. Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A., A hierarchical approach to all-atom protein loop prediction. *Proteins Struct. Funct. Bioinformat.* **2004,** *55* (2), 351-367.

98. Suite 2012: Prime, version 3.1, Schrödinger, LLC, New York, NY, 2012.

99. *Suite 2012: Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012*.

100. Suite 2012: LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2012.

101. Suite 2012: Schrödinger Suite 2012 Induced Fit Docking protocol.

102. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R., Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2005,** *49* (2), 534-553.

103. Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A., New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.* **2006,** *14* (9), 3160-3173.

104. Sherman, W.; Beard, H. S.; Farid, R., Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* **2006,** *67* (1), 83-84.

105. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I., PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2003,** *36* (5), 1277-1282.